APPLIED STATISTICS

A Journal of the Royal Statistical Society

EDITORS
DONALD G. BEECH
AND
LEONARD H. C. TIPPETT

ASSISTANT EDITOR:
BERNARD T. RAMM

VOLUME FIVE

Oliver and Boyd Ltd.
London: 39A Welbeck Street, W.1
Edinburgh: Tweeddale Court

1956

21+1^60

CONTENTS

No. 1. MARCH

	Page
The Use of Statistics in the Investment of Funds *G. H. Ross Goobey, FIA, FIS*	1
Modified Control Limits .. .. *David Hill*	12
A Regression Control Chart for Use in Personnel Selection *Frank Sandon*	20
Factor Analysis as a Technique for examining Complex Data on Foodstuffs .. .. *Roland Harper*	32
Field Work in Sample Surveys of Household Income and Expenditure *Dorothy E. Cole*	49
Notes and Comments	62
Questions and Answers	63
Industrial Applications Section of the Royal Statistical Society: List and Reports of Meetings	65
Study Section of the Royal Statistical Society: List and Reports of Meetings	67
Book Reviews:	
Chance and Choice by Cardpack and Chessboard, Volume 2, Lancelot Hogben	68
Statistical Methods for Research Workers, Sir Ronald A. Fisher ..	68
An Introduction to Stochastic Processes, M. S. Bartlett	70
Population Statistics and their Compilation, Hugh H. Wolfenden ..	71
Publications Received	72

No. 2. JUNE

Examples of Interviewer Variability taken from Two Sample Surveys *Percy G. Gray*	73
Some Sampling Problems in connection with Accounting Records *Stanley F. James*	86
A Tasting Experiment .. .. .. *Norman Theodore Gridgeman*	106
Are Correlations any Guide to Predictive Value? *Robert Ferber*	113

CONTENTS (*continued*) Pa

A 2^4 Factorial Experiment in Four Blocks of Eight: A Study in Radioactive Decontamination *Martin K. Barnett and Frank C. Mead, Jr.* 12

An Extension of the Use of the χ^2-Test *Erik K. Leander and David J. Finney* 13

Question and Answer 13

Industrial Applications Section of the Royal Statistical Society: List and Reports of Meetings 13

Study Section of the Royal Statistical Society: List and Reports of Meetings 14

Book Reviews:

Stochastic Models for Learning, R. R. Bush and F. Mosteller .. 14

Contributions of Survey Methods to Economics, George Katona, Lawrence R. Klein, John B. Lansing and James N. Morgan 14

The Communication of Ideas, T. Cauter and J. S. Downham .. 14

Handbook of Industrial Engineering and Management, W. G. Ireson and E. L. Grant (Editors) 14

No. 3. NOVEMBER

Foreword .. *E. S. Pearson, CBE, President, Royal Statistical Society* 14

Statistics in Hospital Planning and Design .. *Norman T. J. Bailey* 14

Monte Carlo Methods and Industrial Problems *W. Neil Jessop* 15

The Allocation of Scarce Materials between Products *Noel Williams, FIA, ACWA* 16

Industrial Censuses in Great Britain and the United States *Rita J. Maurice* 17

A Technique for studying the Effects of a Television Broadcast *William A. Belson* 19

Missing Values in Experiments analysed on Automatic Computers *Michael Healy and Michael Westmacott* 20

A Comment on the Construction of Price Index Numbers *K. S. Banerjee* 20

Notes and Comments 21

Meetings of Sections of the Royal Statistical Society 21

Book Reviews:

Prediction Methods in relation to Borstal Training, Hermann Mannheim and Leslie T. Wilkins 21

Correlated Random Normal Deviates, E. C. Fieller, T. Lewis and E. S. Pearson 21

Publications Received 21

THE USE OF STATISTICS IN THE INVESTMENT OF FUNDS

G. H. ROSS GOOBEY, FIA, FIS
Imperial Tobacco Company Limited

This article describes the use of various methods of presentation of statistical data in guiding investment policy.

Sources of Statistics Employed

The sources of the statistical information available are many and varied, and the following is a selection of them.

1. Official Publications

Daily	.. ..	*Stock Exchange Official List.*
Monthly	.. ..	*Stock Exchange Monthly Supplement.*
Yearly	.. ..	*Stock Exchange Year Book.*

2. Financial Papers

Daily	.. ..	*Financial Times*, and the financial pages of such papers as *The Times*, the *Daily Telegraph*, etc.
Weekly	.. ..	*Investors' Chronicle; Stock Exchange Gazette; The Economist; The Statist;* and the financial pages of such papers as the *Sunday Times*, etc.
Fortnightly	.. ..	*Investors' Review.*

3. Statistical Services

The Exchange Telegraph Statistical Service.
Moody's Statistical Service.
Statistical services supplied by various stockbrokers.
Stockbrokers' monthly investment lists.

4. Stock Indices

The Actuaries' Investment Index.
The *Financial Times* Indices (daily and weekly).
Investors' Chronicle Monthly Index.
Moody's Index.

5. Company Reports

6. Special Statistical Investigations by Stockbrokers and Jobbers

e.g. (*a*) Investigation into the effect on company profits by the removal of EPL.

(*b*) Investigations into specific industries (some brokers specialise in certain of these) such as Insurance; Breweries; Tea; Aircraft companies; Shipping and shipbuilding; Investment Trusts, etc.

(*c*) Price trends in various classes of securities over many years, e.g. comparisons of the yield on gilts and equities.

J. I. Mason and D. Sachs in their paper to the Royal Statistical Society, 'The Sources and Nature of Statistical Information in Special Fields of Statistics' (*J.R. Statist. Soc.* A, **118**, 67), give an exhaustive list of the data available on this subject.

The following sections illustrate the application of such statistics in practice.

Framing of Policy

(*a*) *Long-term.* Fig. 1 is a simple illustration of how graphical representation can assist in the framing of investment policy. It compares the price of Consols since 1796 with the price index of Industrial Ordinary shares since 1866. This points out in no uncertain manner

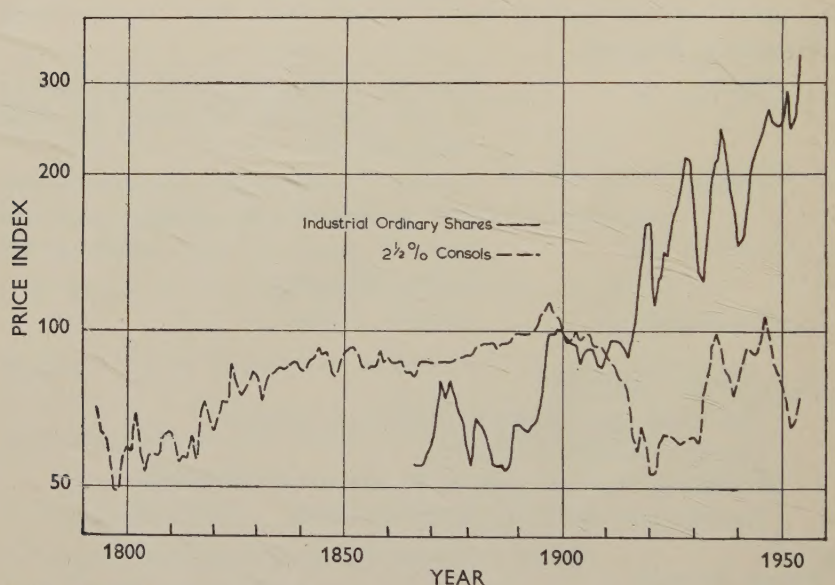


Fig. 1. Variation of price indices of Industrial Ordinary shares and 2½ per cent Consols (Year 1900 ≡ 100 in each case).

the overwhelming long-term advantages of investment in Ordinary shares to an institutional investor whose liabilities are long-term and who is not embarrassed by short-term fluctuations, which are greater with Ordinary shares than with gilts. The object of this paper is not

to discuss all the other considerations which have to be taken into account but to indicate what an important contribution the simple statistical representation of facts can make to these important decisions.

(*b*) *Short-term.* These same graphs (Fig. 1) draw attention to the fact that there are very well-defined cycles, and indicate that in order to get the best possible result an Investment Manager must modify his general policy at certain stages of the price cycle. For instance, when the graphs indicate that one is near the bottom of a trough in the curve, then if one has faith in the future repetition of such trends one would commence buying the sort of Stocks which experience has shown have the greatest possible appreciation when markets are rising. On the other hand, when one is approaching a peak, then one could take refuge in Stocks which are least vulnerable to depreciation, e.g. short-dated British Government securities or, in the limiting case, cash. Of course, one of the most difficult things to do is to decide when one is at the bottom and when at the top, and here it is not so much statistical and graphical considerations which help the Investment Manager, as his own judgement as to what are the causes contributing to the change in trend.

The various national statistics that are published help to fill in the general background picture within which such decisions have to be taken. They help to indicate the general state of the whole economy or of some section of it. In these days of the ever-present fear of a balance-of-payments crisis, for example, the monthly announcement of the country's gold and dollar reserves often has a direct effect on the gilt-edged market.

The Choice of Investments within the Investment Policy

It is obviously most important that when one is considering an investment in a particular company one should have as much information as possible concerning that company readily available in order that one may judge its future prospects. The success or otherwise of an investment depends entirely on the skill with which the Investment Manager can judge the future trend of profits in the particular company. For this purpose he relies largely on the statistical service supplied by the Exchange Telegraph Company or by Moody's. My own Fund uses the Exchange Telegraph service for historic reasons; of course, it is most useful to know one's way quickly over the card to extract the vital points of information. These statistical cards give the important details of the company, its Directors, Authorised and Issued Capital, Debentures, etc., with a summary of the profit and loss accounts and details of dividends for the past ten or so years, followed by the balance sheets for the past two or three years. The method of operation of this service is that a new card is sent out to all subscribers every time there is a change in the facts on the card (e.g. latest profit figures, dividend announcements, etc.). The Moody's service consists of two cards, one containing the main information and a smaller one

containing the information which is most liable to be changed. This enables them to produce a much smaller replacement card on most of the occasions when changes occur. There is, of course, other information which an Investment Manager must have before him when considering a new investment, in addition to the facts on the statistical card. This includes details of present holding in a particular company, for in addition to maintaining a 'spread' between industries it is equally advisable not to have too large an investment in any particular company, for that company through bad management or other circumstances may fall on lean times. Other details are the number of shares offered, the price, and the yield at that price (after allowing for expenses of purchase) on the expected dividend.

Priority Percentages

There are some statistics over which there is a divergence of opinion about the method of presentation. One of these is the statement of the amount by which the earnings cover the present dividend. What one is attempting to assess in considering this figure is by how much there can be a setback in profits before the present dividend paid ceases to be earned. If a company has only Ordinary Share Capital, then if the present dividend declared is one half of the present year's earnings, we know that the net profits can drop by 50 per cent before the company would be forced to decrease the dividend, without eating into the carry forward or the reserves. If, however, there is a Debenture or Preference Capital ranking prior to the Ordinary dividend, it may be that although the present dividend is only 50 per cent of the amount available for the Ordinary dividend, the profits would require to drop by much less than 50 per cent before the dividend is in danger, depending on what proportion of the total earnings the prior charges absorb by way of interest.

As an example, which for the sake of simplicity ignores the question of tax, let us consider a company with Issued Capital of £200 000, of which £100 000 is in 5 per cent Preference Stock and the remaining £100 000 in Ordinary Stock, and assume that the profits are £20 000. The Preference dividend will absorb £5 000, leaving £15 000 available for the Ordinary shareholders. If the dividend declared is 5 per cent (absorbing £5 000 of the £15 000 available for the Ordinary shareholders) one school of thought would describe this as being 'three times covered', thus giving the impression that the profits could drop to one-third of the present level before the dividend is endangered.

A much safer method of presentation is to describe the various distributions and 'carry forwards' as 'priority percentages'; in the case cited above this would be as follows:

	Priority Percentages
Preference Dividend	0– 25
Ordinary Dividend	25– 50
Carry forward	50–100

TABLE I

Analysis of investment holdings according to the sections of the Stock Exchange Daily Official List as at 31st December 1954

Industry	Debentures	% of Fund	Preference	% of Fund	Ordinary	% of Fund	Totals	Total % of Fund
	£ s. d.		£ s. d.		£ s. d.		£ s. d.	
A. Banks and Discount Companies..	—	—	39 526 12 9	0·17	151 644 14 0	0·65	191 171 6 9	0·82
B. Breweries and Distilleries ..	216 268 7 0	0·93	441 124 3 1	1·90	1 056 746 6 8	4·55	1 714 138 16 9	7·38
C. Canals and Docks	2 187 5 10	0·01	10 201 0 0	0·04	39 466 16 0	0·17	51 855 1 10	0·22
D. Commercial, Industrial, etc. ..	1 215 482 16 0	5·23	4 729 013 4 5	20·36	8 786 552 2 2	37·84	14 731 048 2 7	63·43
E. Electric Lighting and Power ..	22 161 17 0	0·10	7 609 2 4	0·03	1 147 4 7	0·00	30 918 3 11	0·13
F. Financial Trusts, Land, etc. ..	296 174 4 10	1·28	98 114 8 8	0·42	275 207 18 6	1·18	669 496 12 0	2·88
G. Gas..	—	—	—	—	21 469 12 9	0·09	21 469 12 9	0·09
H. Insurance	—	—	—	—	140 199 13 3	0·60	140 199 13 3	0·60
I. Investment Trusts	114 840 11 9	0·49	48 144 16 3	0·21	66 579 7 8	0·29	229 564 15 8	0·99
J. Iron, Coal and Steel	61 879 15 10	0·27	187 437 18 0	0·81	1 042 046 2 2	4·48	1 291 363 16 0	5·56
K. Mines	24 565 0 1	0·11	11 635 10 1	0·05	32 867 16 4	0·14	69 068 6 6	0·30
L. Oil..	58 644 18 9	0·25	63 026 10 6	0·27	89 282 16 7	0·38	210 954 5 10	0·90
M. Shipping	7 458 5 3	0·03	52 866 19 4	0·23	31 358 6 4	0·14	91 683 10 11	0·40
N. Tea and Coffee	14 212 10 0	0·06	—	—	3 622 5 9	0·02	17 834 15 9	0·08
O. Telegraphs and Telephones ..	29 705 3 9	0·13	—	—	7 647 10 8	0·03	37 352 14 5	0·16
P. Tramways and Omnibus ..	23 493 5 5	0·10	15 534 1 0	0·07	24 310 4 11	0·10	63 337 11 4	0·27
Q. Waterworks	24 334 7 4	0·10	111 777 1 1	0·48	308 399 18 6	1·34	444 511 6 11	1·92
Total ..	2 111 408 8 10	9·09	5 816 011 7 6	25·04	12 078 548 16 10	52·00	20 005 968 13 2	86·13

Group D is far too great a 'miscellaneous' group for this purpose, and in practice this group is further subdivided into 32 industries.

It is then immediately apparent that the Ordinary dividend is only twice covered, or in other words, profits could only be reduced by one-half before putting the ordinary dividend in danger.

In addition, of course, to the purely statistical facts on the card the Investment Manager endeavours to obtain further information, such as the company's reputation in the trade and up-to-date information as to the company's present prosperity and immediate prospects; but the basic statistical information supplied by these cards is obviously a very material help.

Spread of Investments

I am convinced that in the long run Ordinary shares will prove a much more advantageous form of investment for long-term Funds than the more traditional gilt-edged securities. One of the inherent principles, however, in deciding to invest one's funds in Ordinary Stocks and Shares is to maintain a 'spread' of investments between the various industries, because of the fluctuations which take place from time to time between industry and industry, and which one cannot always foresee. This brings me to another very elementary set of statistics which we maintain in connection with the investments of the Fund with which I am associated. Progressive totals of the amounts invested in the various industries are maintained, and each month these are expressed as percentages of the whole amount invested in order that one may spot at a glance whether the amounts invested in any particular industry are reaching undue proportions. Table I gives a hypothetical example.

Statistics of Ordinary Share Dividends

Another statistical investigation which I have found interesting and useful is set out in Table II, and it may be of interest to recount how this record came to be started.

A Fund which may have embarked on an untraditional investment policy by a large-scale investment in Ordinary Stocks and Shares is naturally particularly interested in satisfying itself that this policy is in fact a profitable one. Now, as part of the routine of keeping all of our investments under surveillance, one of our accountants makes a practice of perusing, as they are sent to us, each of the Annual Reports of the companies in which we have an investment. His job is, of course, to draw attention to anything which appears worthy of investigation if the company concerned has had a bad year, so that the question of the disposal of the investment may be considered, if further investigation should indicate that it might be advisable. Not unnaturally, therefore, the companies with good results are passed over without further investigation, whereas any company which has reduced its profits, or worse still, reduced its dividend, is brought very forcibly to notice. Some time ago this accountant, no doubt because of his approach to this problem, indicated that he was definitely under the impression

TABLE II

Comparison of actual and expected dividends from Ordinary shareholdings of companies whose final dividends are received during the stated months

	Companies Investigated	Dividend Increases	Dividend Decreases	Dividend Unchanged	Aggregated Expected Income	Net Increase	Increase as Percentage of Expected Income
1954:					£ s. d.	£ s. d.	
January	27	14	—	13	25 333 11 6	2 721 10 4	10·74
February	24	11	1	12	18 562 2 9	918 12 3	4·95
March	36	22	1	13	27 219 12 6	4 194 10 0	15·41
April	55	29	2	24	45 918 5 6	5 470 2 7	11·91
May..	60	34	3	23	56 248 8 0	6 351 8 6	11·29
June..	61	33	3	25	61 375 4 8	7 847 2 1	12·79
July	83	50	5	28	68 629 10 10	9 999 12 1	14·57
August	28	15	3	10	29 350 13 8	3 291 1 4	11·21
September	29	11	3	15	28 465 6 3	1 725 12 5	6·06
October	26	16	—	10	29 812 1 10	5 519 5 8	18·51
November	26	19	1	6	31 206 19 8	5 544 15 10	17·77
December	59	35	3	21	67 182 10 9	8 590 3 4	12·79
Total ..	514	289	25	200	489 304 7 11	62 173 16 5	12·71

that of the companies whose accounts he examined, the majori decreased their dividends. This prompted the institution of a monthl investigation of companies declaring final dividends in the previo month, and it showed, as I imagined that it would, that his impressio was entirely wrong. In fact, Table II shows that of the 514 companie from whom we received dividends on Ordinary shares during the pa year, 289 increased their dividend whereas only 25 decreased it. mention this merely as an indication that even the most simple statistic serve to correct wrong impressions. This particular statistical investiga tion goes further than numbers of companies involved and deals also with the percentage increase, and it may be that a graph of these monthly percentage net increases in dividend may be of assistance in indicating whether a trade boom (and therefore a Stock Exchange boom) has come to an end. I rather fear, however, that since dividends are a reflection of trade some many months earlier, the indication will come too late to be of much use in this connection.

A Statistic to Forget

In connection with present holdings, when considering whether or not one should realise an investment because of forebodings as to the future trend of profits of the company, there is one statistic which I maintain should be completely ignored and forgotten. I refer to the price which one has paid for that particular investment. To my mind this price may prove most misleading, but unfortunately, human nature being what it is, it often looms largest in the consideration of the possible sale of an investment. I think that no indication of the future of a company is given by the particular price that one individual paid for shares at a certain date in the past, but how often do we see the rigid policy that a sale will not be considered if the share shows loss on the purchase price, and contrariwise, how often is a share sold which never should be sold just because it shows a profit. It is one of the great misfortunes of accountancy that when a Stock is purchased it is entered in the books at the cost price, which then pursues one throughout the whole ultimate life of the investment.

Gilt-edged 'Switches'

I now come to special cases where the use of simple graphs can assist to produce short-term profits. These are more generally possible in gilt-edged securities where there is no stamp duty involved in an exchange and the percentage commission rates are much smaller generally because amounts dealt in are larger.

By way of illustration, Fig. 2 gives the ratio of the price of 4 per cent Consols over $3\frac{1}{2}$ per cent Conversion. It will be seen that although these two Stocks are similar in character, differences in supply and demand create quite widely fluctuating ratios. When the ratio is high one sells 4 per cent Consols and buys $3\frac{1}{2}$ per cent Conversion, and vice versa when the ratio is low. Typical switches are shown: from £100 00

Consols in May, reversed in July, showing a profit of £2 500; and from £100 000 Conversion in November, reversed in January, for a profit of £3 400.

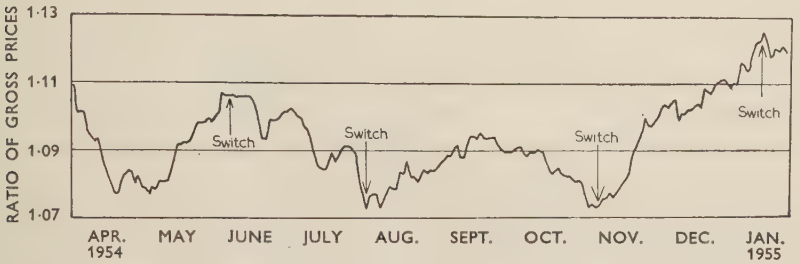


FIG. 2. Ratio of gross prices, 4 per cent Consols/3½ per cent Conversion, showing opportunities for 'switching'.

These graphs are kept by different statisticians in a variety of ways:

1. The prices of two Stocks which it is wished to compare are plotted side by side on the same graph and the price difference observed. The prices plotted can be: (*a*) Gross market prices; (*b*) Market prices less gross accrued interest; (*c*) Market prices less accrued interest less tax at 8s. 6d.; (*d*) Market prices less accrued interest less tax at 7s. 6d. (for the convenience of Insurance companies).

2. The ratio method of our example, using any of the above four variations.

3. Each Stock is similarly compared to a common denominator, such as 2½ per cent Consols.

4. The difference in price of two Stocks is plotted, such prices again being any of the four variations indicated above.

5. Instead of prices, the yields are compared by any of the above methods.

6. Logarithmic scales are often used to give effect to differing price levels.

Switches in Ordinary Shares

These 'graphical' exchanges are possible in other spheres and the shares of the Unilever organisation provide an interesting example of this. This large group of operating companies is owned and controlled by two holding companies—Unilever Ltd, registered in London, and Unilever NV, registered in Holland. There is an agreement between the two companies that all the assets in the Britsh Empire are in Unilever Ltd, the remainder being in Unilever NV. In a winding-up the assets would be pooled. Dividends are declared by each company so

that, allowing for the rate of exchange at the time of payment, they will be approximately the same. Thus, apart from the fact that one is a Dutch and the other a British share, they are virtually identical. Nevertheless, mainly owing to the relative levels of interest rates in London and Amsterdam and other extraneous factors, the prices of these two shares are not normally the same, the difference being sometimes as much as £1. By timing exchanges carefully it is possible to move from one Stock to another while maintaining intact one's interest in the company and to make a capital profit even after allowing for the expenses involved in the exchange.

Another pair of companies with somewhat similar possibilities is Shell Transport and Trading, and Royal Dutch Petroleum. Again we have a British and a Dutch company with shares quoted in London and with common financial interests. Here the position is that the two holding companies own 40 per cent and 60 per cent respectively of each of a large group of operating oil companies generally known as the

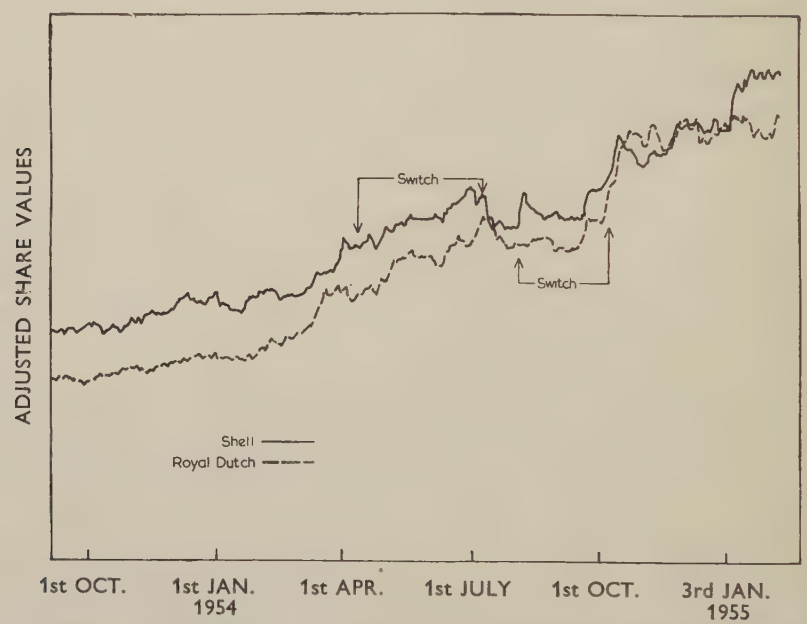


FIG. 3. Adjusted share values of Shell Transport and Trading, and Royal Dutch Petroleum, showing points of advantageous 'switching'.

Royal Dutch/Shell Group. There is no specific equalisation agreement as in Unilever, but obviously as the only assets of each company are the commonly owned operating companies, their fortunes must be very closely tied together. The capital structures of Shell and Royal Dutch are quite different and there is the 2-to-3 sharing of profits to be allowed for, so that a straight comparison of market prices would not be appropriate. However, after making adjustments for the effect of the prior charges and increasing the Shell value by 50 per cent to put

them on to an equivalent footing with Royal Dutch, we plot these adjusted share values on a graph (Fig. 3). There is no intrinsic reason why these two lines should diverge at all, or if they do, it should be at a constant difference. Variations result mainly from the influence of foreign dealing in Royal Dutch, which is a much more international share; the graph shows when one share is relatively dear compared with the other and when a profitable switch is likely.

On the graph are marked two opportunities during 1954 of satisfactory switching, of which advantage was taken. On the 9th April 1954, 4 000 Shell were sold and 500 Royal Dutch were bought, reversing on the 12th July 1954 at a profit of £2 120, or 10 per cent on the capital involved. On the 10th August 1954, 1 100 Shell were sold to re-invest in 150 Royal Dutch. This was reversed on the 7th October at a profit of £600, again 10 per cent.

Many more examples could be given, but I think I have given enough to show how statistics can play a very vital part in the investment of large funds.

MODIFIED CONTROL LIMITS

DAVID HILL
Statistical Advisory Unit, Ministry of Supply

The fixing of quality control limits is not invariably straightforward, particularly when the use of modified limits is possible several methods for calculating modified limits have been described. In this article Mr Hill suggests an approach to the problems that may arise.

THE idea behind modified control limits is that when a process is capable of giving variability about its mean which is only slight compared with the tolerance specified by the designer it is unnecessary to insist upon keeping the mean rigidly stable. Economy in production will be attained by allowing the mean to wander as far as it can safely do.

There are several ways in which such wandering of the mean might occur. It might, for example, follow a steady trend or it might follow a pattern of random variability, or a combination of both trend and random variability. If it is possible to make an investigation of a process to examine the pattern of variability of the mean, much may be learnt about the process and it may be possible to set up a control chart system designed for the particular case so as to combine good control with production economy. It is not, however, always possible or desirable, to carry out such an investigation, and there is certainly a case for a system that will allow control limits to be widened (where it can be done with safety) and that can be applied easily. Modified control limits are designed for this purpose.

The process, of course, must be a suitable one for the job and the wandering of the mean must have some degree of consistency. Obviously, no control chart scheme can be expected to cope with violent and completely unpredictable fluctuations of the mean. If it is assumed that the process is a reasonable one in these two respects, two questions arise: (i) When should natural limits be used, and when modified limits? (ii) Where should the modified limit lines be drawn?

The answer to the first question depends upon the purpose for which the limits are needed. If the idea is to derive the greatest possible information about the product, to produce the most stable product possible, or to seek for causes of instability, there is no doubt that natural limits should be used. If, however, the idea is to produce as many articles as possible to specification as economically as possible then the answer is that the sensible limits to use are whichever are the wider. For if the modified limits are the wider, it means that some variability in the mean can safely be allowed, and it is possible to take advantage of this. If, on the other hand, the natural limits are the

wider, they should be used, for there are disadvantages in using limits narrower than the natural ones. The sample mean can be expected to vary as far as these even if the population mean is completely constant, and, therefore, to use anything narrower will lead to the unrewarding process of searching for 'assignable' causes of the 'unassignable' variation. When the natural limits are the wider, it may be that the process is incapable of meeting the specification. This situation therefore requires careful handling and is mentioned again later.

Which limits will be the wider in any particular case will depend upon the answer to the second question—where should the modified limit lines be drawn? In the use of control limits, as in most other statistical matters, two types of error are possible: the error of taking action when there is no need, and the error of not taking action when there is need. The narrower the control limits compared with the variability of the process the greater is the probability of the first of these errors; the wider the control limits the greater is the probability of the second. Somewhere a balance must be struck.

The classical approach, using natural limits, strikes the balance by keeping the probability of needlessly looking for trouble very low, the argument being that to do so often would soon destroy confidence in the limits as a means of detecting trouble. The probability of failing to detect trouble, when trouble is in fact present, is thus left to look after itself. It can be changed by changing the sample size, but the sample size is usually decided on grounds of convenience, and not with the object of controlling this probability.

There cannot, I think, be much doubt that this philosophy, of considering only the first probability mentioned above, is correct where natural limits are concerned; but unfortunately it has been carried over into the realm of modified limits, where, in my opinion, it is unsuitable.

If the sample size is to be decided in advance on grounds of convenience, we can decide only one of the two probabilities, but with modified limits the one we should choose to be small is, surely, the probability of not taking action when it is needed. For, if it is worth using modified limits at all, it means that we have space to spare within the tolerance range, and we can afford to take action, on occasions, before it is strictly necessary. On the other hand, although we wish to allow the mean to vary as far as it can safely do, we do not wish to miss the chance of a warning when the safety limit is reached. The action taken when a limit is crossed will generally be merely resetting the machine to take the mean away from the limit concerned. When we are using natural limits, to set the mean further from a limit when in fact action is not necessary will produce trouble at the other limit, but when using modified limits we have room to spare, and this will not occur.

The original idea of modified limits was as shown in Fig. 1, the argument being that production is all right (to the extent that only one

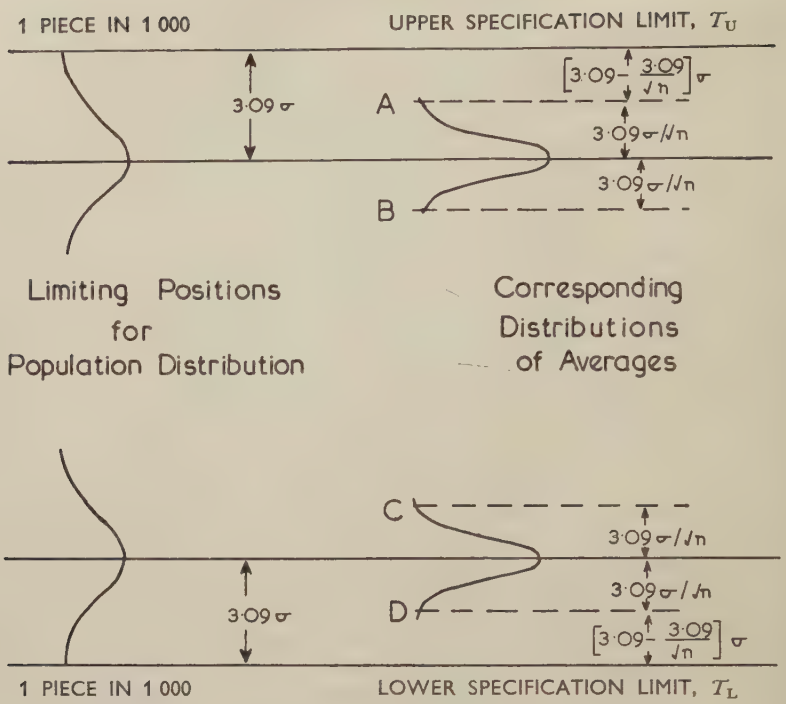


FIG. 1. The derivation of modified control limits, as originally used.

article in a thousand is defective) if the population distribution take the upper of the two positions shown. The natural control limits in thi case would be A and B. Similarly, production is all right if the popula tion takes the lower position—giving control limits C and D. Produc tion is certainly satisfactory anywhere between these limiting position of the distribution. So if we use A and D as our control limits, we shal supposedly keep production satisfactory, and at the same time get th economic advantage of much wider limits than the natural ones.

This argument leads to limits drawn inside the tolerances at a distanc of $(3{\cdot}09 - 3{\cdot}09/\sqrt{n})\,\sigma$, where n is the sample size in use and σ is th standard deviation of the process. (This distance is estimated by $M\bar{i}$ when mean range, $\bar{w}$, is used instead of standard deviation. The facto M can be found in *A First Guide to Quality Control for Engineers*[8] and i given also in Table I of this article.)

The fallacies of this argument are: (i) Even though production ca nominally be considered satisfactory at either of the extreme positions if truly in control, an article at the extreme of the tolerance is unlikel to be as satisfactory as one near the centre. (ii) When a process ha shown itself to be in control, a point within the control limits can b taken as an indication that it is likely to be still in control; if a poin is close to, say, the upper limit it is a reasonable assumption that th population mean is unchanged, and is lower than the observed point But if a process is not in control, in the sense of complete stability—an

the whole point of modified limits is that complete stability is not enforced—then there are no reasonable grounds for such an assumption. If a sample mean is near the upper limit, it is still true that the population mean may be well below it, but equally true that it may be well above it! (iii) It has been forgotten that the positions of the natural control limits are not perfect from every point of view. They have been adopted as the best compromise to solve a particular problem. There is no logical reason for assuming that they are reasonable for a quite different problem.

Another reason for caution is that normality is being assumed. When natural limits are used, normality is assumed in the distribution of the mean. This is a reasonable assumption, since means tend to normality, even when the population from which they are drawn is not normal. But the modified limits assume normality of the population as well—a much more doubtful assumption.

These limits are (except for the use of the factor 3·09 instead of 3) those given in a series of articles by H. Rissik[6] in 1943 and later in his book *Quality Control in Production.*[7] He appears to have had no doubts on their applicability, but other writers have been more cautious. Thus E. H. Sealy in *A First Guide to Quality Control for Engineers*[8] takes the limits further from the tolerances, using $(3{\cdot}09 - 1{\cdot}96/\sqrt{n})\sigma$ (or $K\bar{w}$ where mean range is used) and saying: 'When modified control limits are being used in place of the theoretically correct limits . . ., it is found necessary to allow some margin of safety. . . . We therefore make use of the "inner limits" of *B.S.* 600R.' He also gives the caution: 'One word of warning when using modified control limits. They are to some extent a concession on the true theoretical form of control, and this concession should not be abused. Keep, therefore, comfortably within the modified limits and never allow runs of points just inside them. Arrest tool-wear drifts, if possible, well before they actually reach the limits.'

This advice, in effect, leads to the calculation of limits which are then not used. Instead use is made of alternative limits that are not drawn on the chart, are arbitrarily narrower, and are adopted on the basis of the intuition of the user. If narrower limits are necessary, it should be possible to give them, on a probability basis, in such a form that they can be used as they stand.

M. J. Moroney, in the first edition of *Facts from Figures*,[4] in his example of modified limits uses limits $(3 - 3/\sqrt{n})\sigma$ inwards from tolerances but cuts his tolerance range to two-thirds of the design value as a safety factor, before calculating his limits. Again, if such a safety factor is necessary, and in my opinion it is, then it should be devised upon a probability basis, not upon a purely arbitrary basis. For then it will be known that the safety factor is large enough to give the desired safety, but nevertheless it will be no larger than necessary, thereby giving the maximum economic advantage from the use of modified limits.

In *Engineering Statistics and Quality Control*[1] I. W. Burr discusses the

limits $(3 - 3/\sqrt{n})\sigma$ and $(3 - 2/\sqrt{n})\sigma$ inwards from tolerances, a considers the latter to be the better of the two. He also states that t maximum and minimum safe process average lines (given by limits inwards from tolerances) should always be drawn on the chart as we and that these are the fundamental things to watch.

In *B.S.* 2564[3] Dudding and Jennett give six different modified limit based upon $(3{\cdot}09 - 3{\cdot}09/\sqrt{n})\sigma$; $(3{\cdot}09 - 1{\cdot}96/\sqrt{n})\sigma$; $(2{\cdot}33 - 3{\cdot}09/\sqrt{n})\sigma$ $(2{\cdot}33 - 1{\cdot}96/\sqrt{n})\sigma$; $(4 - 3{\cdot}09/\sqrt{n})\sigma$; and $(4 - 1{\cdot}96/\sqrt{n})\sigma$.

The first two of these are intended for general use together, givin outer limits and inner limits. The third and fourth give outer an inner limits when the bulk is permitted to contain a small proportio of rejects. The last two are given as alternative modified limits t provide higher assurance when required. All these limits are base upon the probability of taking action when it is not required, but wit different percentages defective considered to be just satisfactory. Th advice given is similar to Sealy's, namely: 'When using these modifie control limits it is desirable to take action when sample average approach a limit.'

L. H. C. Tippett, in *Technological Applications of Statistics*,[9] having pointed out that 'it is logically fallacious to transfer all the argument that apply when there is statistical control to the uncontrolled situation' asks the question: 'Why are the modified control limits set outside th limits for the population mean rather than inside?'

I believe that there would be advantages in setting them inside, an suggest that useful limits would be obtained by placing them so tha when production reaches the state of producing one part in a thousan defective there is only one chance in twenty of *not* taking actio Assuming a normal population such limits would be drawn inside th tolerances at a distance given by $(3{\cdot}09 + 1{\cdot}645/\sqrt{n})\sigma$. These limi may be used just as they are. There is no need to take action until sample mean is beyond a limit; there is no need for safety factors. Th safety factor is built into the limits. This factor is sufficient to b reasonably safe even if the population is not in fact normal.

It will be noticed that these limits, unlike most control limits, a wider for a large sample size than for a small one. It is surely righ that they should be. The usual modified limits accept a sample mea nearer to the tolerance when there is less information, than when ther is more information—clearly an undesirable feature. The usua modified limits, if a point on the chart is just *less* than $3{\cdot}09\sigma$ from a tolerance, adopt the attitude: 'I am now probably making defective pieces, but the evidence is not yet definite enough for me to be sure so I shall carry on for the time being,' On the other hand, the limit now proposed, if a point on the chart is just *further* than $3{\cdot}09\sigma$ from a tolerance, adopt the attitude: 'I am probably not yet producing defectives, but there is a chance that I may be, so, for safety's sake, had better reset the machine. The tolerances are easy on this job, s there would be no real excuse for abusing them.'

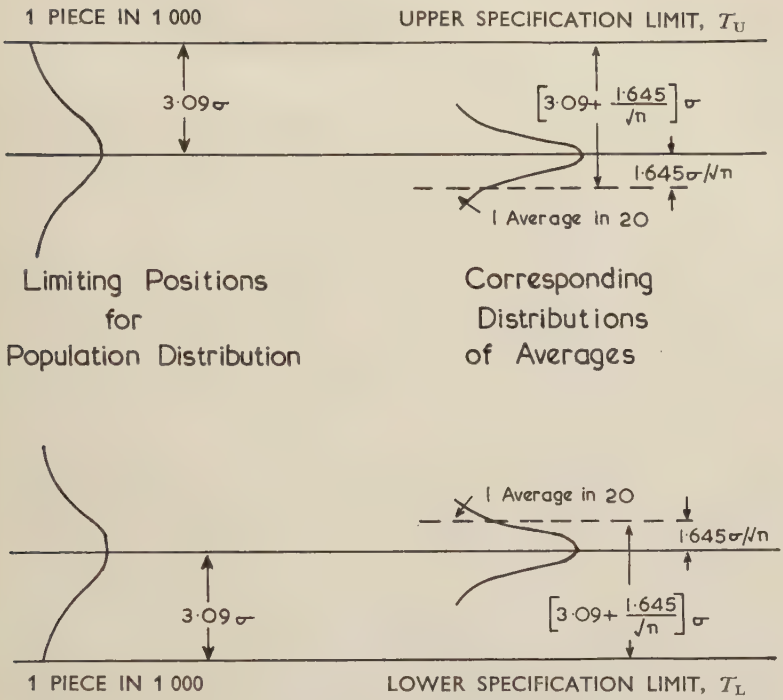


FIG. 2. The derivation of the modified control limits proposed in this article.

These proposed limits are shown diagrammatically in Fig. 2, and may be derived by multiplying the mean range of a number of samples ($\bar{w}$) by the factor $\mathcal{J}$ given in Table I. The result is the distance to draw the limits inwards from the tolerances. This factor $\mathcal{J}$ is calculated as $(3{\cdot}09 + 1{\cdot}645/\sqrt{n})a_n$, where a_n is the factor used in estimating a standard deviation from a mean range of samples of size n.

TABLE I

Control chart constants

Sample Size (n)	L	H	M	$\mathcal{J}$
2	0·18	0·09	0·80	3·77
3	0·27	0·15	0·77	2·39
4	0·33	0·19	0·75	1·90
5	0·38	0·22	0·73	1·64
6	0·41	0·25	0·72	1·48
7	0·44	0·28	0·71	1·37
8	0·46	0·30	0·70	1·29
9	0·48	0·32	0·69	1·22
10	0·50	0·34	0·69	1·17

If (Mean Range)/(Drawing Tolerance) is greater than L, use natural limits and 100 per cent sorting in addition.

If between L and H, use natural limits, checking that neither limit is nearer to a tolerance than $M\times$ (Mean Range).

If less than H, use modified limits, drawn $\mathcal{J}\times$ (Mean Range) inwards from Drawing Limits

If the upper limit is denoted by T_U and the lower limit by T_L, the width of the tolerance is $(T_U - T_L)$ and the width of the modified limits as now proposed will be $(T_U - T_L) - 2(3{\cdot}09 + 1{\cdot}645/\sqrt{n})\sigma$ whereas the width of natural limits is $6{\cdot}18\sigma/\sqrt{n}$. Natural limits should, therefore, be used whenever $6{\cdot}18\sigma/\sqrt{n}$ is greater than $(T_U - T_L) - 2(3{\cdot}09 + 1{\cdot}645/\sqrt{n})\sigma$; i.e. whenever $\bar{w}/(T_U - T_L)$ is greater than H, where H is given in Table I, and is calculated as $1/a_n(6{\cdot}18 + 9{\cdot}47/\sqrt{n})$. It is worth using these proposed modified limits whenever $\bar{w}/(T_U - T_L)$ is less than H.

As long as action is taken whenever either the mean chart or the range chart shows a point beyond the limit, these modified limits ensure a product almost entirely within tolerance. When natural limits are in use, however, merely to keep the charts in control may not ensure good product. If $\bar{w}/(T_U - T_L)$ is greater than L (given in Table I and calculated as $1/6{\cdot}18a_n$) a perfect product is not attainable. A control chart with natural limits is of use in keeping the scrap as low as possible, but 100 per cent sorting is necessary as well.

If $\bar{w}/(T_U - T_L)$ lies between L and H, natural limits are capable of producing a good product provided strict control is kept with the population mean in the correct place. (Checking that neither limit is nearer to a tolerance than $M\bar{w}$ is necessary to make sure that the mean is within the permissible range.) The fact that these limits can produce a good product, although they are wider than the modified ones, is not a contradiction. For if the product is kept within control using natural limits, there is evidence of stability, and it is quite reasonable that a particular point on the chart may be considered safe when there is evidence of stability, although it would be considered as a warning of possible danger in the absence of such evidence.

In the second edition of *Facts from Figures*[5] Moroney considers a different approach to modified limits, designed to allow the choosing of both the probability of taking action when it is not necessary, and the probability of not taking action when it is necessary. If both these have been chosen the sample size cannot be chosen as well; this has to be calculated to give the desired probabilities. The same idea is given by R. Cavé in *Le Contrôle Statistique des Fabrications*,[2] where an *abac* is given for deriving the sample size and the positions of the control limits.

This is an interesting approach but, to my mind, not a very profitable one, for it takes away from the man who wishes to operate control charts both the right to choose his sample size, and the simplicity of calculation for the fixing of the control limits. It gives him very little in return, for although he now has the right to choose what fraction defective he considers bad quality, what fraction defective he considers passably good quality, and the two probabilities of making the wrong decisions, he will rarely have other than an arbitrary basis on which to make these choices.

Furthermore, if this approach is adopted, the fraction defective chosen as representing passably good product, for which action should

not be taken, is likely to be very small. This means that we are concerned with a point some way out in the tail of the frequency distribution and so the assumption of a normal distribution is important. If the population is not normal the calculated probabilities may bear little relation to the facts, in spite of Moroney's claim that 'these charts can be used with every confidence, even where the distribution of the individual items is extremely skew.' Another difficulty is that, if the sample size cannot be chosen, but must be calculated, it may turn out to be inconveniently large—in the example Moroney discusses it is thirteen. It does appear to be a common human characteristic that, when attention is diverted from the sample size to the probabilities, the probabilities of taking wrong decisions are often chosen to be needlessly small. This is likely to lead to comparatively large samples, and is a disadvantage for two reasons.

The first is that the time and labour used upon a large sample may be less acceptable than the use of a small sample size with the rather more frequent resetting that the limits proposed here might imply. The second reason is more important. If the sample size derived is rather large, what is to be done about a range chart? As the sample size increases, the range becomes less and less efficient as a measure of spread, and is not to be recommended for samples larger than ten or so. Yet in using modified limits a range chart is very necessary, for the whole procedure of letting the mean vary as far as it can do with safety assumes the variability to be constant and known. It is important to keep a strict watch on range to check that the variability in fact remains constant at this known value. The alternative of a standard deviation chart would, again, usually be less acceptable than rather more frequent resetting. There is, of course, the possibility of drawing a large sample, using all the measurements to calculate the mean, but only some of the measurements, selected at random, to calculate the range. Such devices, however, add complication and mean that information is being used inefficiently.

REFERENCES

[1] Burr, I. W. (1953). *Engineering Statistics and Quality Control.* McGraw-Hill, New York and London.

[2] Cavé, R. (1953). *Le Contrôle Statistique des Fabrications.* Editions Eyrolles, Paris.

[3] Dudding, B. P., and Jennett, W. J. (1955). *B.S.* 2564, *Quality Control Technique when Manufacturing to a Specification.* British Standards Institution, London.

[4] Moroney, M. J. (1951). *Facts from Figures* (1st Ed.). Penguin Books, Harmondsworth.

[5] *Idem.* (1953). *Facts from Figures* (2nd Ed.). Penguin Books, Harmondsworth.

[6] Rissik, H. (1943). *Aircraft Engineering*, **15**, 55, 85 and 115.

[7] *Idem.* (1947). *Quality Control in Production.* Pitman, London.

[8] Sealy, E. H. (1945). *A First Guide to Quality Control for Engineers.* HMSO, London.

[9] Tippett, L. H. C. (1952). *Technological Applications of Statistics.* Williams and Norgate, London.

A REGRESSION CONTROL CHART FOR USE IN PERSONNEL SELECTION*

FRANK SANDON

Selection Officer, Birmingham Education Committee

Candidates from different schools may be characterised by two sets of data: 'scores' and 'marks'. Scores, being based on teachers' assessments, reliably measure the relative qualities of the candidates from any one school, but the standard of scoring may vary from school to school, so that scores for candidates from different schools are not directly comparable. Marks are derived from the results of a common examination; they are not as reliable as scores for any one school but are comparable for different schools. Mr Sandon shows how, with the aid of correlation analysis, marks may be used to adjust scores so that scores for all candidates can be made comparable.

Introduction

Suppose that a school sends a number of its boys for consideration for a limited number of posts within a large organisation. The head of the school has had the lads under personal observation for a number of years, and he is able to rank them in an order of suitability for the job. Not merely this, but he can assign 'scores' to all his boys on, of course, his own 'local' standards, these scores therefore showing not merely the rank but the relative sizes of the 'gaps' between the successive boys. On his own standards these scores can be regarded to be as sound a measure as can be achieved in educational work. The boys are probably not a random sample of all his pupils.

The personnel officer of the organisation, who has to make the selections for the posts, thus has this batch of boys from school A. He also has batches of boys from schools B, C, D, etc. For each batch he has a set of scores for the boys from the separate heads, but the standards of the individual heads differ, and the various sets of scores cannot be directly compared. To make the best choice of the candidates before him the personnel officer commonly in such cases makes all the candidates work a simple examination, meet a selection board, or submit themselves to an interview. Whatever the procedure, from it he gets a 'mark' which will be a second numerical measure of each boy's suitability.

* This article is based on lectures given to the Midland Branch of the British Psychological Society on 11th December 1952 and to the Birmingham and District Group of the Industrial Applications Section of the Royal Statistical Society on 9th December 1953.

The personnel officer realises that this latter measure, like any other examination mark, is subject to large errors, and is not at all a reliable measure of the relative merits of the candidates; but he wishes to make the best use of his rather inaccurate mark, together with the scores which, however accurate within schools, are not on any objective standard. What can he do? The problem has been met with in a number of educational selection cases, and several solutions have been suggested. None of them, as far as I am aware, has recognised both the errors arising from small samples and also the probable accuracy of the estimates of the relative 'gaps' between the boys that their schools have made.

In all such instances it would be in accordance with general experience that there would be a correlation, though possibly only a small one, between the score and the mark for each school, and the method outlined below is an attempt to make use of this fact. The number of candidates from each school is small, very frequently less than one dozen, and sampling errors are large. But there will be certain schools for which the regression line will be so far away from that for the great majority of schools that we shall be justified in saying that it is most likely that the 'local' standard of scores for that school is at fault, and that the head has scored his boys on a level that is too high or too low. In such a case adjustment of the school's score is called for in order to get the boys from that school fairly fitted in with the great bulk of the candidates.

The suggested method possesses certain features in common with the method used by Mr Desmond for investigating rating in time study,[1] but it has not, as far as I know, been used by any other worker for dealing with the problem in the form under discussion.

The numerical examples that follow are taken from a selection procedure in which the candidates were pupils who were the subject of reports from their individual head teachers and who were given a common allocation examination. The examination, being a one-day 'sudden death' affair, was, as suggested above, not considered to give such a clear picture of the boys as the reports given by the schools in which they had been under daily observation for several years.

Score and Mark for Candidates from One School

For a particular school presenting eight boy candidates the figures given in Table I were obtained. The correlation coefficient between score and mark—rather meaningless in view of the truncated nature of the distribution of the scores, lads from the tail of the distribution not being presented as candidates—is 0·596, and the regression of score on mark in working units is 0·292. These values are satisfactory considering the crudeness of the mark and the fact that the two measures are not even estimates of exactly the same features of the various boys —the mark was on a written test, the score was on general suitability for the job. From the figures we derive the regression line to give the

TABLE I

Mark and score for eight boys of school B

Candidate	Mark	x*	Score	y*
a	646	14	122·9	21
b	644	14	122·9	21
c	680	18	122·5	21
d	682	18	119·9	19
e	653	15	117·8	18
f	693	19	117·4	18
g	636	13	113·5	16
h	548	4	110·9	15
All eight	$\Sigma x = 115$ $\Sigma x^2 = 1\,811$ $\bar{x} = 14{\cdot}375$		$\Sigma y = 149$ $\Sigma y^2 = 2\,813$ $\bar{y} = 18{\cdot}625$	
	Correlation coefficient = 0·596 Regression coefficient, y on x = 0·292			

*x and y are in working units, both to the nearest unit:
$x = (\text{mark} - 504{\cdot}5)/10$; $y = (\text{score} - 81)/2$

best estimate of the score for any known mark to be, in terms of actual units:

$$(\text{score} - 118{\cdot}25) = (0{\cdot}058\,4)\ (\text{mark} - 648{\cdot}25)$$

This line is shown in Fig. 1.

The corrected sum of squares of y, i.e. the sum of squares of the scores about the mean score, arises from two 'causes'. Part of it is due to the random fluctuations that arise from the fact that the repeated measures for the score of any one boy, or of the various measures of the score for a number of boys obtaining the same mark, would not be always exactly the same. Part of it is due to the fact that as the mark x increases so does y, the score. The contribution of the latter part is the sum of squares due to regression and can be estimated, in working units (using the data of Table II for school B), as $(46{\cdot}13)^2/157{\cdot}9$, i.e. as 13·5. The balance of the sum of squares, here 37·9 — 13·5 = 24·4, is therefore the sum of squares not accounted for by regression. It is due to the difference between the way the teacher sees each boy and the way the boy performs in the test or, alternatively, to the different ways the teacher might see boys who get the same mark. These differences arise in part from different points of view and in part from the teacher's 'personal equation'. This residual sum of squares, 24·4, divided by the number of degrees of freedom, here two fewer than the number of boys, gives us 4·07 as the variance (in working units) not accounted for by regression. The square root of this, 2·016, multiplied by two to bring to actual units, is 4·032 and is the best estimate of the standard deviation of the way in which, all things considered, the school will range its

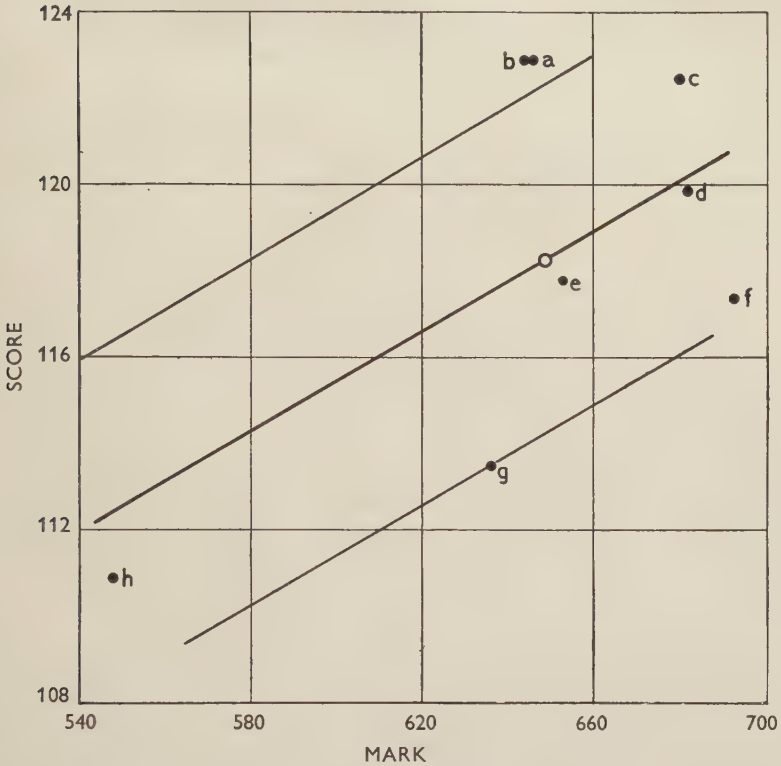


FIG. 1. Regression line of score on mark for school B, with parallel lines at a distance of one standard deviation about the regression.

estimates of score for boys having the same mark. In Fig. 1 are drawn two parallel lines, each one standard deviation away from the regression line. Included between them are practically all the observations.

In connection with the effects of sampling, we must remember that not merely the score but also the mark may be affected, in that for each boy this is based on only a limited sample of his attainments, etc., and in that he may not have been at the top of his form when he worked the test.

Relation between Score and Mark for Different Reporters

We have now found the regression line to give the score for any prescribed mark in the case of a particular school, and we also know how closely to it will lie the estimates made by the school for the scores appropriate to any one mark. Do all schools estimate similarly? As suggested above, we must be prepared for differences between the standards of the various reporters—in our case the heads of the different schools. Let us look at these differences and see if they are greater than could be due to such mere sampling effects as would arise if reporters, all with exactly the same standard, were scoring, each his own small set of boys.

TABLE II

Regression calculations for nine schools (working units as for Table I)

School	n	$\bar{x}$	$\bar{y}$	$\Sigma(x-\bar{x})^2$	$\Sigma(y-\bar{y})^2$	$\Sigma(x-\bar{x})\times(y-\bar{y})$	Regression Coefficient	Residual SSy	d.f.	Residual Variance
A	12	18·9	14·8	232·9	90·3	87·75	0·377	57·2	10	5·72
B	8	14·4	18·6	157 9	37·9	46·13	0·292	24·4	6	4·067
C	15	14·4	10·3	595·6	231·3	323·00	0·542	56·2	13	4·323
D	11	15·8	8·8	201·6	143·6	116·56	0·578	76·2	9	8·467
E	14	15·4	12·2	141·4	396·4	3·71	0·026	396·3	12	33·025
F	9	15·1	12·1	69·0	152·9	90·89	1·317	33·2	7	4·743
G	9	14·4	11·0	120·2	284·0	152·00	1·265	91·7	7	13·10
H	10	15·4	17·5	46·4	36·5	−7·00	− 0·151	35·4	8	4·425
I	13	14·5	13·0	251·2	230·0	111·00	0·442	180·9	11	16·445
Total	101	—	—	1 816·2	1 602·9	924·04	—	951·5	83	11·464
All schools	n	X	Y	$\Sigma'(x-X)^2$	$\Sigma'(y-Y)^2$	$\Sigma'(x-X)\times(y-Y)$	Regression Coefficient	Residual SSy	—	—
	101	15·4	12·9	2 008·5	2 444·6	991·99	0·494	1 954·7	99	—

Σ is summation over a school, Σ' is summation over all boys, $\bar{x}$ and $\bar{y}$ are school means, X and Y are grand means, SSy means Sum of Squares of y.
For the regression line with common slope going through school means:

Regression of y on $x = 924{\cdot}04/1\,816{\cdot}2 = 0{\cdot}509$

SSy due to this $= (924{\cdot}04)^2/1\,816{\cdot}2 = 470{\cdot}3$

Table II gives the means and the regression coefficients for nine schools. The lines are neither parallel nor concurrent. We could expect them to differ slightly in slope, or to pass within a slight distance of a common point; this would be the result of sampling variations. But the differences may be due to more than sampling effects—they may be due to differences in the standards and in the discriminating powers of the various reporters. We shall examine these possibilities, following the methods given by Tippett.[2]

Do Reporters score at the Same Level?

We test this by calculating the residual sums of squares of y for deviations from (i) the common regression line for the 101 boys, going through the grand mean of x and y, and from (ii) regression lines having a common slope but going through separate means for the nine schools. The sum of squares for residual (i) is 1 954·7, and the degrees of freedom are $101 - 2 = 99$. The sum of squares for residual (ii) is $1\,602{\cdot}9 - 470{\cdot}3 = 1\,132{\cdot}6$, and there are $101 - 10 = 91$ degrees of freedom.

For eight extra degrees of freedom, lines (ii) reduce the sum of squares by an additional $1\,954{\cdot}7 - 1\,132{\cdot}6 = 822{\cdot}1$, as compared with line (i). The mean square is 102·76. This may be compared with the residual mean square of $1\,132{\cdot}6/91 = 12{\cdot}4$ based on 91 degrees of freedom. The ratio of mean squares is $F = 8{\cdot}3$, and for 8 and 91 degrees of freedom this is overwhelmingly significant.[3, 4, 5] Thus we conclude that the nine schools give scores at different levels for any given sets of marks.

Do Reporters spread their Scores Equally?

Now we test whether the regression coefficients are significantly different. We do this by calculating the sum of squares of residuals from (ii) regression lines having a common slope but going through separate means for the nine schools, and from (iii) separate regression lines for the nine schools, each going through its own appropriate mean. The sum of squares for residual (ii) is, as above, 1 132·6 and the degrees of freedom are 91. The sum of squares for residual (iii) is, from Table II, $57{\cdot}2 + 24{\cdot}4 + \ldots + 180{\cdot}9 = 951{\cdot}5$, and there are $101 - 18 = 83$ degrees of freedom, giving a mean square of 11·5.

For the difference between (ii) and (iii) the sum of squares is $1\,132{\cdot}6 - 951{\cdot}5 = 181{\cdot}1$ and the degrees of freedom are 8, giving a mean square of 22·7. The ratio of the mean squares is $F = 2{\cdot}0$, and for 8 and 83 degrees of freedom this corresponds to a probability of 0·06. This is near the lower conventional level of significance.

Do the Scores for all Reporters fit in similarly with the Marks?

Reporters may differ among themselves in yet a third way. They might have worked to the same general level and might have produced the same regression line. Yet they might have correlated their scores

with the personnel officer's marks to different degrees. As we have pointed out above, we cannot usefully use any values for correlation here because the distribution of scores for any school will be curtailed. We must therefore adopt an alternative approach. The most suitable is that of a test for homogeneity of residual variances. There are two main tests for this that are available, both of them using a comparison of the weighted geometric mean with the arithmetic mean of the variances of the separate schools. Of the two available, the more convenient is Bartlett's test.[2, 3]

We have from Table II that 83 times the logarithm (to base 10) of the geometric mean of the variances is $= 10 \times (\log 5{\cdot}72) + 6 \times (\log 4{\cdot}067) + \ldots + 11 \times (\log 16{\cdot}445) = 77{\cdot}156\,1$. The logarithm of the arithmetic mean of variances is $\log (11{\cdot}464) = 1{\cdot}059\,3$. Subtract from $83 \times 1{\cdot}059\,3$ the value 77·156 1, to give 10·765 8. Multiply this by 2·302 6, the factor to translate logarithms to base 10 to natural logarithms. This gives 24·789 3. Then Bartlett tests for this by taking it as a χ^2 with 8 degrees of freedom (there are nine schools involved). Hence P is about 0·01, and we conclude that the schools are not giving scores that fit the marks all in the same way.

But we notice that School E seems to be contributing an undue share to the sum of squares. Let us test with this school omitted. The sum of the degrees of freedom is 71, the sum of the residual sums of squares is 555·2, and the weighted sum of the logarithms of the variance of the individual schools is 58·930 5. We thus test for a χ^2 equal to $(2{\cdot}302\,6) \times [71 \log (555{\cdot}2/71) - 58{\cdot}930\,5]$, i.e. 10·331 1, for which P lies between 0·5 and 0·1. Greater accuracy in applying the test demands that we should use a correcting factor for inequality of size of schools, which in our case is 1·04. This does not bring the value of χ^2 to a significant level.

We conclude, therefore, that there is no significant heterogeneity in the residual variances of the eight schools: the teacher of School E had some exceptional aspects of his lads in mind when he marked them.

Application of the Results

We have now applied three tests to the relationships between marks and scores for the different schools. Although there are marked differences between schools in respect of the discrimination effected by their scores, and although there are marked differences between schools in respect of the closeness with which the scores fit the marks, yet the overwhelmingly significant difference between schools is in respect of the levels at which they score.

It may be true that the connection between mark and score on which alone we can base any scaling of scores for the different schools is very slight. Every method that has been suggested for dealing with the problem has been forced to assume that there is some connection. We shall make the same assumption. We shall at the same time adopt as a working assumption that on the whole the relative discriminations

assigned by the different schools are much the same. We shall turn to deal with the gross differences in level that we have found to be so very significant. How then can we best correct for departures from the standard so as to make the scores of the different schools more fairly comparable? We shall not, in this article, consider further the question of attempting to correct for the other differences.

Since we are assuming that the slope of the regression of score on marks to be the same for all schools, we need only concern ourselves with the school averages. These, in actual units, can be derived from the values in working units of Table II. We have:

School	A	B	C	D	E	F	G	H	I	All
Mean mark	692	646	647	661	657	654	647	657	648	657
Mean score	110·5	118·3	101·7	98·6	105·4	105·2	103·0	116·0	107·0	106·8

The points are plotted in Fig. 2. The point representing the grand mean for all schools is shown by a small circle.

The mean scores cannot be compared directly, since they correspond to different mean marks. But we can make an allowance for these latter by drawing a regression line through the point representing the grand mean with a slope representing the relationship between mark and score when the standard of assessment is unchanged. The best estimate that we have of this is the regression coefficient, 0·509 in working units (Table II). In actual units of mark and score this is 0·102. The resulting regression line is thus

$$(\text{score} - 106{\cdot}8) = 0{\cdot}102\ (\text{mark} - 657)$$

and is drawn in Fig. 2. Departures from the average level of scoring are measured by the deviations, in score, of points from this line. According to this criterion, School A is very near the average.

In order to correct scores we subtract for each school the deviation, measured along the y-axis, of its point from the line. For School B, for example, the value to be subtracted is

$$118{\cdot}3 - 106{\cdot}8 - 0{\cdot}102\ (646 - 657) = 12{\cdot}4$$

For School C it is $101{\cdot}7 - 106{\cdot}8 - 0{\cdot}102\ (647 - 657) = -4{\cdot}1$.

A Regression Control Chart

In applying the above procedure it is well to recognise that some of the deviations from the line may be due to the fact that there are only a few boys from each school. The school means are thus subject to error. We shall ignore such deviations and ignore corrections that are small compared with their errors.

These errors are best estimated from the mean residual variance from the lines that have a common slope but that go through the individual school means. This variance, in working units, is 1 132·6/91

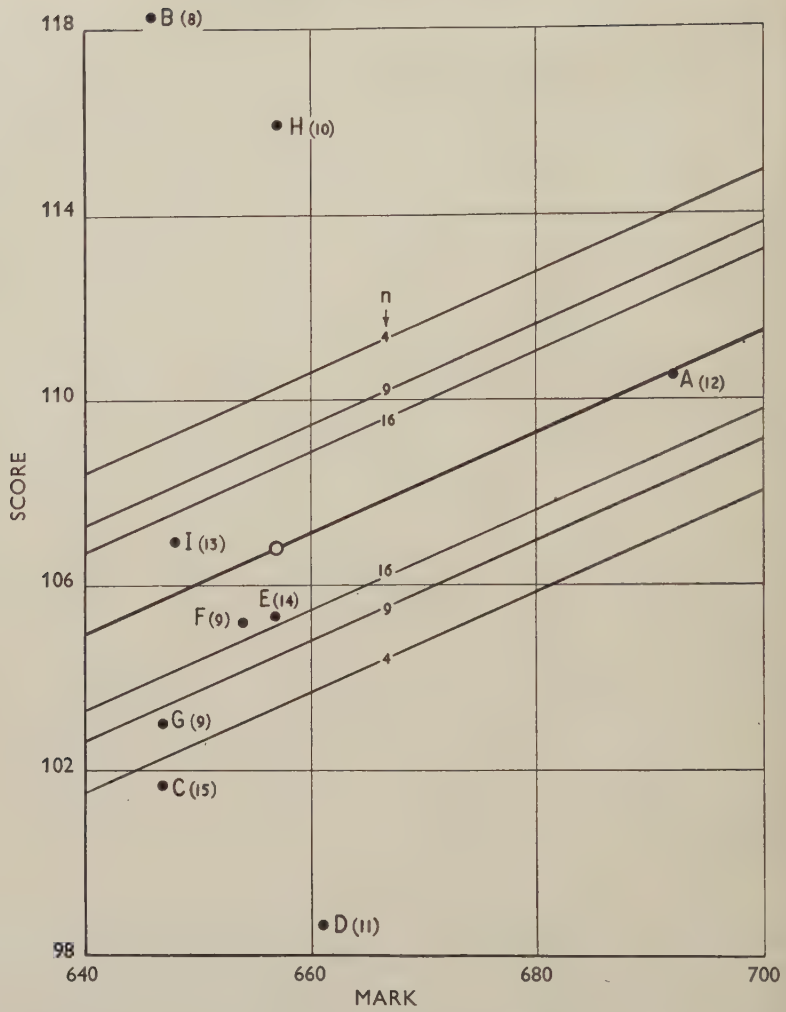


FIG. 2. Control chart for mean score against mean mark applicable to different values of the number of boys from a given school. (Calculated from the data in Table II.)

= 12·45, giving a standard deviation of 3·53, or 7·06 units of scor If the number of boys from a school is n the standard error of the mea score is $7{\cdot}06/\sqrt{n}$, and we may draw on the chart of Fig. 2 control line parallel to the regression line but distant from it by this amount. Ther if there are no real differences between schools, approximately two thirds of the points will lie between these lines. We arbitrarily regar any school with a point outside the lines as being atypical to a degre requiring the application of a correction to its scores, and apply th correction developed in the previous section. No correction is made fo schools with points between the lines.

The value of n differs for different schools, so we have drawn, o Fig. 2, sets of limit lines for several values of n. The point for eac school must be assessed with reference to the appropriate pair of lim

lines—that for School C, for example, with reference to lines for $n = 15$, which are just beyond those drawn for $n = 16$. According to the chosen criterion, correction to the scores would be made only for schools B, C, D, G, and H.

Other Sampling Effects

In the above cases we used, as the basis for the determination of the regression line and of the control lines, the data from nine schools. The schools were not a true random sample of all the schools concerned. They were all schools of average 'size'. The reason for choosing them was that in this way we avoided using very large schools, which gave heavy computations to produce information about one school only, and avoided also very small schools, which, by the time we had lost two degrees of freedom per school, did not give much information. The nine schools chosen gave a fair number of schools and a fair number of candidates without unduly increasing the labour of the calculations and of the checking. It may be noted that in practice we have found no feature about rating peculiarities that is common to, say, the large schools or to, say, the small schools.

It should, however, be remarked that the above procedure takes no account of sampling errors in the determination of the grand means, the regression coefficient, and the residual variance. These differences will not affect substantially our solution, as we wish to reduce all schools to what is an approximate common basis, and no question will arise (though corrections will be more numerous) if the common basis is faulty. But, more important, any correction that may be applied will itself be subject to a considerable random error unless the number of boys in the school is large.

The control lines have been put at distances of once times the standard error of the mean, to give two-thirds of cases within each pair of such limit lines. The usual convention with control charts is, of course, to take twice the standard error (to include 95 per cent of the cases), or three times the standard error (to include 399 of 400 cases). We suggest that once times the standard error is preferable for this work, because with the wider control lines the determination of a big correction or none at all may depend on a very small difference in the position of the mean. On the other hand, we cannot pinch much closer than one standard error, or the limit would be no deviation at all. This would mean that we would bring every case from a school with one boy only down on to the regression line, and thus we would virtually refuse to accept any report and value of score from such a school.

Unfortunately, however, by placing the control lines so close to the regression line as we have done we may have a fair proportion of cases where the rating is really on the common basis but where the points are misplaced because we have a truly exceptional sample from the cases so rated. The result is that a correction may unnecessarily, though not, it is submitted, unjustifiably, be made.

I see no practical way out of this difficulty, or of the other difficul
that arise through the sampling effects just mentioned above, th
would be theoretically satisfactory. I am not proposing a perfe
accurate procedure. This is never possible or even meaningful in a
case of measurement in education and here, as so very often, I mer
propose to make the best of the data available.

A Convenient Procedure

The use of a graphical method, of the regression line and of t
control lines, is a device to meet the practical needs of scaling individu
reports about boys from a large number of schools. These reports a
often submitted on individual file sheets, and these cannot always b
conveniently handled to give the means per school. There is an alte
native method that has been found to be of service in actual use. Th
consists of plotting the individual points and estimating the position
the school means. It can be done rapidly as follows.

Draw, once and for all, the regression line through the grand mea
and a convenient number of pairs of control lines. Plot for each bo
his mark and score by placing on the chart a small coloured butto
Do this for all the boys from one reporter. Take a button of a differe
colour and place it at the estimated position of the centre of gravity
the points. This would be, for example, at the midpoint between tw
at the intersection of the medians of the triangle formed by three, a
for more than three at other positions that can with practice be est
mated with a fair degree of accuracy. The position of this point relati
to the appropriate control line is then obvious, and the correction,
thus determined to be necessary, is at once found by the deviation
the score at this point from the regression line.

The Regression Control Procedure Summarised

In summarising the method developed above I shall use the ter
'mark', 'score', 'boy', and 'teacher' for simplicity though, as point
out earlier, the procedure is not to be limited to such cases.

(*a*) Obtain for each boy the school's score and the personnel officer
(unreliable) mark.

(*b*) Take a sample of schools that will give estimates of the mean
mark and of score, the regression of score on mark, and the residu
standard deviation (the variation not accounted for by such regression

(*c*) Plot the regression line and pairs of control lines for suitable 'size
of school.

(*d*) Plot, for each school in turn, the mark and the score of each bo

(*e*) If the mean score of any school lies outside the control line for th
appropriate size of school adjust the score to bring the mean on to th
regression line, correcting each score by a flat number of units
score.

(*f*) Accept other cases, and use the unadjusted scores.

(*g*) Make offers in accordance with the scores, adjusted or unadjusted, thus obtained.

I wish to point out that the responsibility for the statements, methods and conclusions of the above paper are my own and not in any way those of the Birmingham Education Committee.

REFERENCES

[1] Desmond, D. J. (1951). 'The statistical approach to time study', *Statistical Method in Industrial Production.* Royal Statistical Society, London.

[2] Tippett, L. H. C. (1952). *The Methods of Statistics.* 4th edition. Williams and Norgate, London; Wiley, New York.

[3] Brownlee, K. A. (1949). *Industrial Experimentation.* 4th edition. HMSO, London.

[4] Comrie, L. J. (1949). *Chambers Four Figure Mathematical Tables.* Chambers, London and Edinburgh.

[5] Lindley, D. V. and Miller, J. C. P. (1953). *Cambridge Elementary Statistical Tables.* Cambridge University Press.

FACTOR ANALYSIS AS A TECHNIQUE FOR EXAMINING COMPLEX DATA ON FOODSTUFFS

ROLAND HARPER
Department of Psychology, University of Leeds

One of the problems arising in the appraisal of foodstu concerns the complex relationships that may exist betwe 'objective' tests and subjective judgements. In this article I Harper discusses the potentialities and limitations of Fact Analysis in dealing with this problem.

Introduction

In a recent article in this journal I discussed certain of the funda mental problems involved in the subjective appraisal of foodstuffs The observations made were not primarily concerned with any on particular food product, although interest in the whole set of problem arose out of a number of years' work on the skills of professional chees makers and cheese graders. The question was raised as to how on should examine a set of 'complex multi-dimensional relationships' suc as may exist between 'objective' test data and skilled judgements on number of cheeses. In the particular investigations to be considered th objective tests were primarily mechanical in nature, but such da might equally well have included information derived from chemic or bacteriological tests. In the previous paper it was suggested that *o* method of dealing with the particular type of problem which has ju been posed is the technique of Factor Analysis. A brief outline will b given of the main features of this technique, and its use with foo products will be illustrated by these studies of cheeses. It is of intere to add that very recently Baker[1] has used factor analysis to examine th connection between data derived from various forms of routine chem cal analysis and data derived from the evaluation of the quality samples of wine by normal sensory methods. This particular stud involved less than twenty samples of wine, whereas almost two hundre cheeses were used in the studies which will be summarised and discusse later.

The main function of the present article is to examine both th potentialities and the limitations of factor analysis as applied to dat concerning foodstuffs. No systematic examination along these lines ha yet been made. Ostle and Tischer[10] in their otherwise very compre hensive review of statistical methods in food research make no referenc whatever to factor analysis. Yet it is seven years since the first pre liminary study employing data on cheeses was published (Harper an

Baron[5]). In what follows an attempt will be made to present the highly technical concepts involved in factor analysis in a form that will be intelligible to those without direct experience of this technique. However, before the aims and functions of factor analysis are reviewed a few observations should be made concerning the origin of its application to cheeses. The investigations that will be used to represent the technique in action were carried out on behalf of the National Institute for Research in Dairying. They owe their origin to the work of Dr G. W. Scott Blair, in whose department they were carried out. In addition to his pioneer work which led to the development of the various instrumental tests employed, Scott Blair was the first to recognise that factor analysis might lead to a clearer interpretation of the data being accumulated. Specific reference should also be made to Miss Margaret Baron, who was concerned with many different aspects of the field studies. I was concerned mainly with the initial planning and with the final analysis of the data. Five years have now elapsed since I was directly connected with these investigations. Thus it is possible for me to approach the task of evaluating this work with a degree of detachment which would not be possible in the heat of front-line research work.

The Technique of Factor Analysis

This section is concerned with a brief statement in general terms of the aims and the principles underlying factor analysis. It follows along lines already set out in the last chapter of *Foodstuffs, their Plasticity, Fluidity and Consistency*, edited by Scott Blair.[11] Factor analysis should not be confused with factorial design. The latter refers to a systematically planned approach leading to the provision of information about the relative importance of various conditions or processes and the way in which these may interact upon some quantifiable characteristic. There is no danger of the professional statistician confusing these two terms, but they are sufficiently similar for the layman to do so. Many professional statisticians are highly critical of factor analysis, especially on account of its lack of mathematical rigour at certain points. (See Kendall and Babington Smith[8] for an expert discussion on this theme.) One of the special functions of factor analysis is that of providing a method of sorting out the underlying order in a vast amount of empirical information obtained by different methods in the same general situation. Systematically planned studies require that the main variables of importance are already known. In an exploratory stage this state of affairs has not been reached, and factor analysis may sometimes be employed to isolate the main variables. In this exploratory phase, which would correctly describe the present state of knowledge in connection with the systematic assessment of foodstuffs, numerous complex issues are involved. In consequence, minor defects in the mathematical rigour of techniques of analysis may be of only secondary importance. Tools should be related to the task in hand, and unrestricted demands

for academic precision may be out of place in handling certain comp problems of practical significance. While I admit that factor anal has its limitations, which will emerge during the discussion of practi examples, our immediate aim is to achieve general understanding.

As already indicated, factor analysis is employed in an attempt fathom the underlying relationships which may exist between a su stantial number of 'measures' obtained by testing a number of sampl It is assumed here that several different types of tests are involved. principle, the samples tested may be almost any class of objects events. Factor analysis has been elaborated mainly by a number mathematically minded psychologists; in their own investigations t samples tested have usually been people and the tests employed ha been predominantly tests of human performance or mental functionin Eysenck[3] and Vincent[13] have already discussed certain aspects factor analysis in this journal, but some recapitulation of the ma features is necessary. One convenient method of expressing the relatio between two tests is to calculate the correlation coefficient between t 'measures' obtained by applying these tests to the same set of spec mens. Thus if we denote the two sets of measures by x and y respe tively, and make certain assumptions which cannot be elaborate here, the correlation coefficient (r_{xy}) between test X and test Y is give by the expression

$$r_{xy} = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\{\Sigma(x - \bar{x})^2 \cdot \Sigma(y - \bar{y})^2\}^{\frac{1}{2}}} \quad \ldots\ldots(1)$$

In this expression $\bar{x}$ and $\bar{y}$ represent the averages of the two sets measures. The numerator in the expression for r_{xy} is simply the su (over all samples) of the products of the deviations of the x-measure a the y-measure for each sample from the mean values of x and y, respe tively, for the whole set of samples. The denominator takes account the range of variability in the measures. The value of the correlati coefficient calculated in this way will lie between $+1$ and -1. Plus would denote a perfect linear relationship between the two sets measures with x and y increasing together. Minus 1 would still indica a perfect linear relationship, but this time as one set of measur increased the other would decrease in value. Intermediate values of t correlation coefficient denote that the particular measures are impe fectly related. A value of zero would be taken to indicate that r relationship existed. According to the number of specimens tested, t correlation coefficient must exceed a certain value in order to 'significant'. This simply means that a certain range of values includi zero could arise through the normal workings of 'chance' from two se of measures which are 'really' unrelated. For more detailed inform tion the reader who is not well informed on this topic should consu one of the standard introductory textbooks on statistics.

The procedure outlined for calculating the correlation coefficient between two sets of measures can obviously be applied more or less mechanically to measures derived from a substantial number of tests applied to the same set of specimens. This results in a table of correlation coefficients which is sometimes referred to as a *correlation matrix*. Represented in symbols this would take the form:

Test / Test	a	b	c	d	...	...
a	—	r_{ba}	r_{ca}	r_{da}	...	...
b	r_{ab}	—	r_{cb}	r_{db}	...	...
c	r_{ac}	r_{bc}	—	r_{dc}	...	...
d	r_{ad}	r_{bd}	r_{cd}	—	...	...
...	...	...	...	...	—	...
...	...	...	...	...	...	—

The diagonal cells have been left blank because they are not normally determined by direct observation. Terms placed symmetrically with respect to the blank diagonal are, of course, identical, but for the sake of completeness they have been represented in a consistent manner; thus:

$$r_{ab} = r_{ba}, \text{ etc.}$$

For certain purposes considerable economy in space is achieved by expressing the original data in this way, although some information is inevitably lost in the process of condensation. Even twenty tests will yield a table containing 190 independent entries. Such a table of numbers may still be too large to grasp and to interpret effectively. Is it possible to produce an even more condensed statement which will represent the underlying order in the data which have been collected? Let us take another purely hypothetical example also expressed in algebraic terms. By means of the computational techniques of factor analysis it is possible to reduce the so-called correlation matrix to a smaller table, referred to as the *factor matrix*. This assumes that there exists a certain degree of order in the data collected. Let us assume that in the present instance this underlying order can be accounted for by three independent, hypothetical 'factors'. Since we are not at present concerned with the problems of interpretation it is unnecessary to discuss here the precise meaning to be given to these factors. Under the conditions laid down the factor matrix would take the following form:

Test \ Factor	I	II	III
a	a_1	a_2	a_3
b	b_1	b_2	b_3
c	c_1	c_2	c_3
d	d_1	d_2	d_3
...	...	...	...
...	...	...	...

The entries in the table can be interpreted in a number of differer ways. For instance, *within certain limits of error*, governed partly by th point at which one stops 'extracting further factors', it is possible t reconstruct the original correlation coefficients by a series of equation of the type

$$r_{ab} = a_1b_1 + a_2b_2 + a_3b_3 \qquad \ldots.(2$$

(assuming that the data do not justify going beyond factor III).

Up to and including instances in which three factors are required is possible to represent the information contained in the factor matr by graphical methods or by means of solid models. Beyond this lev of complexity some understanding of the geometry of n-dimensions ar some acquaintance with matrix algebra seem desirable. In th examples to be considered it will fortunately be unnecessary to g beyond what can be expressed in terms of three dimensions.

Let us consider two further hypothetical examples before turning t real data collected in the field. It is important to bear in mind that th usual methods of factor analysis can lead only to the isolation of wha are referred to as 'common factors'. Several tests are essential for th effective 'isolation' of one single common factor. The example take for graphical representation is one involving two factors (I and II Attention will be focused upon two tests, a and b, although in practic more than two would be essential to provide the information require to construct these diagrams. Let us consider the manner in whic factor analysis allows us to represent the relationship between thes tests under two rather different sets of conditions.

Case 1. Representation of the relation between two errorless tes involving two, and only two, common factors.

The entries in the factor matrix corresponding to tests a and b ma be treated as if they were the co-ordinates of the points a and b, a illustrated in the two-dimensional diagram given in Fig. 1. The poin a and b have been joined to the centre or origin, and the radiatin lines may be referred to as 'test vectors'. A circumscribing circle

unit radius has also been drawn. This is intended to represent a state of affairs in which any point on the circumference would indicate that the whole of the variation (variance) of that set of measures is completely accounted for by the two common factors. In the present example these are labelled I and II, although it will be evident that the relationship between tests a and b would be preserved even if the reference factors or reference axes were allowed to 'rotate' to new positions. Although such rotations of axes are carried out by some investigators they have not been employed in the present studies. Hence no further reference to this aspect of factor analysis will be made here. It is customary to denote the length of the respective test vectors (in this case unity) by the symbols h_a and h_b. The square of this length is referred to as the 'communality of the test' and indicates the proportion of the test variance accounted for in terms of the common factors which have been isolated. The variance of a set of n measures denoted by the values of x is simply $\Sigma(x - \bar{x})^2/(n-1)$. When more than two common factors are required the full expression for the communality would be:

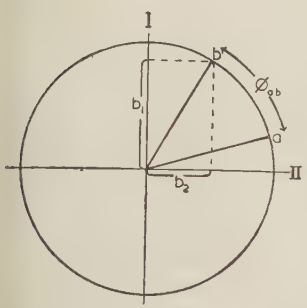


FIG. 1. Representation of the relation between two errorless tests, a and b, assumed to involve only two common factors. (Reproduced from Fig. 13, section 4.1, of ch. VIII of *Foodstuffs, their Plasticity, Fluidity and Consistency*.)

$$(h_a)^2 = (a_1)^2 + (a_2)^2 + (a_3)^2, \text{ etc.} \qquad \ldots.(3)$$

a_1, a_2, a_3, etc., being the successive entries in the factor matrix under the headings of factors I, II, III, etc., continued as far as the computational procedures make it necessary. Although in the hypothetical example given in Fig. 1 h_a and h_b are both unity, this will by no means always be the case. Let us examine this point in relation to Case 2.

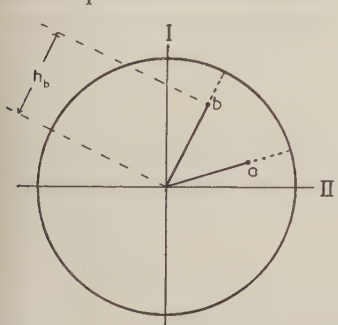


FIG. 2. Representation of the relation between two tests, a and b, involving error and/or specific variance, and assumed to involve only two common factors. (Reproduced from Fig. 14, section 4.1, of ch. VIII of *Foodstuffs, their Plasticity, Fluidity and Consistency*.)

Case 2. Representation of the approximate relation between two tests requiring only two common factors.

In this case it is assumed either that the tests involve errors of measurement or that a certain degree of individuality in the tests makes it impossible to account for all the variance in terms of the two hypothetical common factors. This state of affairs is represented in Fig. 2. In this case both h_a and h_b are less than unity. Whereas in Case 1 the correlation between the two tests could be represented solely in terms of the angle between the two test vectors, in Case 2 it is necessary to take into account the

lengths of the test vectors. In trigonometrical terms the correla
coefficient—or rather an estimated value based upon the com
factors—is given by the equation

$$r_{ab} = h_a h_b \cos \phi_{ab} \quad \ldots$$

The same relationship holds good when more than two dimensions
involved. When this happens the angle ϕ_{ab} is an angle out in spa
If more than three common factors are necessary this becomes an an
in a generalised space of many dimensions. It will be obvious that t
correlation between two tests, ranging as it does between + 1 and −
lends itself to representation in angular form. It could be argued th
the techniques of factor analysis enable one to erect a structure with
which the relationships between the various tests can be expresse
economically in this manner. It is not possible to examine here th
geometrical implications of the various alternative approaches to th
general class of problems which have been stated. However, facto
analysis is related to a number of the more conventional statistica
techniques. A number of important comments on this theme have bee
made quite incidentally by Cronbach and Gleser.[2]

It will be evident that certain of the relationships which underlie
number of sets of measures obtained by employing a variety of tests o
a set of specimens can be revealed and represented within the fram
work of factor analysis. No one would wish to claim that this is the on
way of revealing such relationships. Nor can it be claimed that facto
analysis is capable of dealing with all possible forms of relationshi
The relationships which are revealed are not free from such accident
influences as those due to the selection of a *particular* group of tests f
probing a situation or a *particular* group of specimens to which these tes
are applied. Perhaps it would be better to employ the more gener
phrase 'methods of probing a situation' rather than the more restricte
word 'tests'. Provided that the information can be expressed in quant
tative terms, some indication of the many underlying relationshi
should emerge from the use of factor analysis. Obviously the huma
senses represent one very important method of probing the wor
around us, and many examples of this type provide an opportunity fo
combining objective and subjective data into the same general analysi

Factor Analysis of Data from Cheshire Cheeses

The origin of the investigations to be discussed has already bee
indicated. Many of the early studies by Scott Blair and his colleague
employed correlation methods. These included the well-known tech
nique of 'partial correlation', whereby it is possible to calculate wha
the correlation coefficient between two tests would be if the effects o
one or more other tests affecting the relationship were held constan
Although this process can be repeated to give high-order partial corre
lation coefficients, these are extremely difficult to interpret. It appeare
that factor analysis might provide a more meaningful overall pictur

of the relationship between the various tests being employed than was being achieved by the methods of partial correlation. It may be noted that there is a simple relation between the technique of calculating partial correlation coefficients and the technique of extracting factors according to the rules of factor analysis. Although the majority of the earlier cheese studies focused attention primarily upon instrumentally measurable properties, the extension of the test data to include appropriate subjective assessments was simply a further logical development.

Full reports have already been published of the studies of Cheshire cheeses (see Harper and Baron[6] for a convenient account of this work). Here it is possible only to present the minimum amount of information essential to subsequent discussion. The field work was carried out in

TABLE I

Identifying Letter	Description of Test or Assessment	Essential Details	High Values correspond to:
A	Ball compressor	A measure of surface hardness	Soft, easily indented
– B	Ball compressor	Recovery after removal of load	Small amount of recovery
– D	'Meyer' ball compressor	Measure of 'hardening'	No appreciable hardening
E	Borer (rind removed)	Resistance to penetration	Soft, easily penetrated
G	Needle penetrometer (rind removed)	,,	,,
I	Needle penetrometer (through the rind)	,,	,,
– K	Spring skewer (through the rind)	,,	,,
– L	Breaker	Resistance to fracture	Low breaking resistance
M	Firmness (outer surface)	Resistance to pressure of the thumb	Soft
N	Springiness (outer surface)	Recovery on removing pressure	No springiness
O	Firmness (internal sample)	As for M	Soft
P	Springiness (internal sample)	As for N	No springiness
Q	Crumbliness (internal sample)	Breaking down sample with finger and thumb	Lacks crumbliness
R	Overall quality (internal sample)	Includes taste and smell as well as all other characteristics	Poor quality

Negative signs for B, D, K, and L are consistent with those used in previous reports and with the signs used in Table II.

a commercial cheese factory. The data discussed refer to tests
roughly two hundred cheeses at a time when they would normally h
been marketed. The collection of data for the two studies, which w
carried out in successive years, was spread over about eight we
During this time five cheeses from each day's production were subjec
to eight mechanical tests and six types of subjective assessment. T
details of the various tests and assessments employed are set out
Table I.

The instrumental tests (A–L) were concerned with measuring re
vant complex physical characteristics. These included surface har
ness, resistance to penetration, recovery, hardening as opposed
hardness, and resistance to fracture. Such measurements are releva
for a number of different reasons. Certain of the tests approxima
fairly closely to the manipulative actions of the professional grad
when he samples and grades cheeses. The mechanical behaviour of
cheese is relevant here because there is an intimate relation betwe
these characteristics and the balance between moisture content an
acidity in the ripening cheese. The subjective assessments (M–R) to
the form of simple ratings which involve assigning each cheese to o
of five or seven categories intended to represent the range of probab
variation under the heading of the particular quality or chara
teristic being assessed. Although three members of the grading pan
were involved here, for the purposes of the present study the individu
ratings were simply averaged. In the light of the advanced discussi
set forward in the recent contribution to this journal the subjecti
techniques employed here may appear rather crude (see Harper
However, it must be remembered that the more sophisticated stan
point had not been reached when the present studies were planned a
that the introduction of any systematic form of subjective assessme
into these investigations was a significant advance upon the design
previous work. The task in hand was to make the most effective u
of methods of subjective appraisal which were already to hand rath
than to pursue intensive studies of this important facet of the to
problem. The assessment of firmness and of springiness was made
pressing the thumb into the surface of each cheese. Such assessmer
were first made on the outside of the cheese and then upon a cylindric
sample taken from within by means of a grader's sampling iron. Ev
this action provides the skilled grader with important informatio
about the cheese concerned. The crumbliness of the sample so with
drawn was assessed by breaking down a portion between finger an
thumb. Finally, the assessment of overall quality was made. Th
deliberately took into account taste and smell as well as the chara
teristics which had previously been assessed.

It is impossible to outline here the various computational stages
factor analysis. Eventually, product moment correlation coefficien
were calculated between all pairs of variables, the subjective an
the objective data being treated in the same manner. The 'centro

TABLE II

Centroid analyses of subjective and objective tests on Cheshire cheeses. (Factor loadings with decimal point omitted, e.g. + 85 represents + 0·85)

Factor	I		II		III		IV		V		h^2	
Test	1949	1948	1949	1948	1949	1948	1949	1948	1949	1948	1949 (5 factors)	1948 (4 factors)
E	+ 85	+ 88	+ 08	+ 09	+ 38	+ 16	+ 21	− 21	− 04	—	916	853
G	+ 71	+ 67	+ 40	+ 60	+ 31	+ 23	+ 31	− 10	− 03	—	865	867
Q	+ 19	+ 27	+ 82	+ 75	+ 16	+ 15	− 24	− 03	+ 20	—	828	663
O	+ 85	+ 84	+ 22	+ 25	− 23	− 10	− 09	+ 25	− 25	—	896	846
P	+ 84	+ 67	+ 21	+ 24	− 30	− 34	− 09	+ 19	− 14	—	866	653
R	+ 60	+ 55	+ 40	+ 51	− 33	− 06	− 10	− 15	+ 24	—	692	587
− K	+ 69	+ 66	− 09	− 14	+ 29	+ 25	− 18	+ 26	− 15	—	631	585
I	+ 75	+ 76	− 34	− 31	+ 28	+ 12	+ 06	− 11	+ 06	—	758	704
− D	+ 65	+ 44	− 51	− 51	+ 17	+ 11	− 21	− 20	− 04	—	748	507
A	+ 95	+ 90	− 23	− 26	+ 13	− 10	+ 04	− 17	− 08	—	976	914
M	+ 92	+ 86	− 14	− 18	− 12	+ 21	+ 16	+ 22	+ 11	·	924	867
N	+ 90	+ 74	− 15	− 25	− 18	+ 12	+ 10	+ 12	+ 14	—	888	642
− B	+ 82	+ 62	− 44	− 68	− 20	− 28	− 12	− 21	+ 05	—	923	961
− L	+ 64	+ 39	− 24	− 15	− 38	− 40	+ 15	+ 09	− 17	—	660	346
Variance accounted for (per cent)	58·2	47·1	12·9	16·7	6·9	4·5	2·7	3·1	2·0	—	82·7	71·4

Data modified from:

HARPER, R. and BARON, M. (1951). *Brit. J. Appl. Phys.*, **2**, 35.

BARON, M. and HARPER, R. (1950–1). *Dairy Industries*, **15**, 407; **16**, 45.

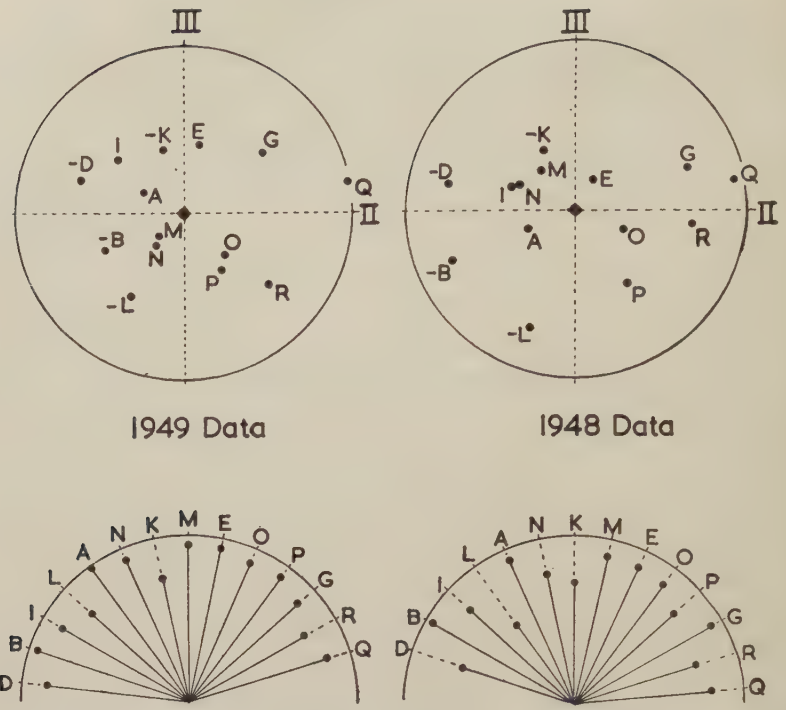


Fig. 3. Representation of the relations between various tests and assessments of Cheshire cheeses, as indicated by factor analysis. (See text for full explanation.)

method of factor analysis', as described by Thurstone[12] (pp. 149–75), was used to 'reduce' the original table of correlation coefficients to a table of factor loadings. Table II gives full details of the factor loadings relevant to the two separate studies. It is not necessary to assign meaning to the various factors at this stage. As in the previous hypothetical examples, the factors represent a set of axes at right angles to one another which will be employed to demonstrate the underlying relationships between the various tests. In the present instance there is a considerable degree of correspondence in the two successive years in so far as the first two factors are concerned and a certain measure of correspondence with respect to the third factor. Hence it is possible to represent the salient features of these two separate studies by means of solid models involving three dimensions. Although such models have been made, mere photographs do not seem to convey effectively the information contained in the model itself. Fig. 3 represents an attempt to convey by means of diagrams what the solid model looks like.

Essentially, the figures given in the first three columns of Table II have been used to lay out a set of radiating vectors corresponding to the various tests or assessments. Not only may these vectors vary in angular position but also in length. In the upper half of Fig. 3 the angular relationship is represented by indicating upon a circular plan

the points where the various test vectors, if produced, would emerge from the surface of the surrounding unit sphere. It will be recalled that a test vector of unit length implies that the entire variation (variance) of that particular set of measurements can be accounted for in terms of the common factors which have been isolated. The additional information about the lengths of the various vectors is presented in a fanwise manner in the lower half of Fig. 3. Any tests that are represented by a vector of less than unit length (i.e. the full radius of the sphere) must satisfy one of several conditions. Either the measurements concerned are subject to error or the observed variation in the particular set of measurements must be attributable to factor(s) which are not represented here. The second condition might arise when an errorless test samples characteristics not covered by any other test or assessment in the group. If no 'overlapping' test can be found to account for this residual component of variation it would be referred to as 'specific variance'.

Let us now consider the main points, which are well represented by Fig. 3.

(i) Considering that the two studies have been carried out in successive years on a material which is subject to biological variability, the general correspondence is surprisingly good. Certainly there are some minor shifts in the positions and lengths of certain test vectors, and these may require interpretation. However, it might be more correct to consider the data as referring to two different populations rather than to two different samples of the same population of cheeses.

(ii) These two analyses force us to recognise that at least 'three dimensions' are necessary to describe the way in which Cheshire cheeses differ from one another. Of course, this type of multi-dimensional variation is common enough in other fields (cf. hue, brightness, and saturation in the field of colour). In the present instance, partly because of the computational techniques, the particular dimensions represented by the factors I, II, and III are not all of the same importance. As already indicated, there is no need to pin our faith on these accidentally determined factors. It is possible either to 'rotate' these axes to a more meaningful set of positions or to pick out specially favoured tests which would effectively provide the complete range of information represented in the three-dimensional picture.

(iii) The next task is, then, to decide upon the best method of representing these three essential dimensions. There is much to recommend the choice of tests which occupy a unique position in the solid model, e.g. Q, — D, and — L. There is much to recommend the choice of tests involving small errors of measurement. Test A satisfied this condition, for the length of the appropriate test vector is almost unity in both studies. There is also much to recommend the selection of uncorrelated tests so that each will provide information not provided by others. Let us start with test A. In addition to involving a small

error component this test comes very close to the position of the 'firs centroid factor' by means of which a very substantial part of the tota variance in all the tests or assessments can be accounted for. Almos at right angles to test A (i.e. largely uncorrelated with it) is test Q which is also in a unique position in the system as a whole. Th identity of these two tests may now be revealed. Test A is an instru mental test of surface hardness which closely simulates the grader' action of 'thumbing' a cheese. Test Q is the subjective assessment o crumbliness, a characteristic for which no adequate instrumenta substitute had been invented, at least up to 1950. If we accept thes two tests with their traditional basis as adequately representing two o the main dimensions, how can we represent the third? If we insis that the third dimension shall be strictly at right angles to (i.e. inde pendent of) the other two it will be evident that no such test exists. A vector at right angles to tests A and Q would occupy a region in which no actual test vectors are to be found. If any definite statement can be made about the nature of such a hypothetical dimension it is tha this must be regarded as a composite variable. All that can be said with any measure of certainty is that this third dimension takes into account the differentiation between variables (— D, I, — K and E) and variables (— L and P). Obviously we are moving away from clearly defined interpretation to ambiguous information which seems to offer a whole range of alternative interpretations. The main distinc tion between these two groups of variables is that between relatively long-time and relatively short-time deformation characteristics. (I this statement is obscure try manipulating 'bouncing putty'—a syn thetic compound which behaves very differently under different condi tions of testing.) However, this is little more than a hypothesis which emerges from the analysis, and which should be submitted to more rigorous methods of testing.

(iv) A few remarks are necessary about the position of the Overal Quality Variable (R). In the 1949 data the vector representing test R lies closer to the third dimension than any other vector. Unfortunately in this series of studies no attempt was made to secure a separate assessment of flavour. Overall quality is what psychologists would describe as a 'global assessment'. Wearmouth and Baron,[14] in extend ing the lines of work which have been outlined to Cheddar cheeses deliberately included a separate assessment of flavour. Subsequen analysis showed that this showed higher correlation with overal quality than with any other test or assessment. It would be satisfying to suggest that the missing third dimension could best be represented by the assessment of flavour if one were forced to choose an actual test However, this is a matter for further investigation.

(v) The reader may enquire why tests Q, — D, and one somewhere near — L have not been proposed as a group for defining and quanti fying the most important features of a Cheshire cheese in view of the

unique positions involved. Reference to the lower half of Fig. 3 will soon make it clear that the lengths of the vectors representing tests — L and — D fall far short of the desirable unit length, and there is reason to believe on several grounds that these are not the most efficient tests, no matter how advantageous their selection might be from a qualitative point of view.

(vi) One direct comparison should be made between the Cheshire cheese studies with those subsequently made on Cheddar cheeses. The majority of tests and assessments were common to both studies. The grading panel was deliberately asked to make separate assessments of both firmness and springiness in the Cheshire studies in spite of the fact that trade experts stated that these two characteristics could not be distinguished. On the other hand, Cheddar cheeses possess a quality known in the trade as 'fight-back'. This might be described as 'the way in which a cheese pushes against your thumb when the pressure has been released'. These traditionally recognised distinctions are fully confirmed by the analyses. Table III gives comparative information about the same group of four tests; details which are not relevant to the purpose in hand are omitted.

TABLE III

Composition of firmness and springiness of Cheshire and Cheddar cheeses. (Factor loadings with decimal points omitted)

	Cheshire Cheese[6]			Cheddar Cheese[14]		
Test \ Factor	I	II	III	I	II	III
—	—	—	—	—	—	—
—	—	—	—	—	—	—
M	+ 92	− 14	− 12	+ 73	+ 19	+ 32
N	+ 90	− 15	− 18	− 37	+ 09	− 26
—	—	—	—	—	—	—
—	—	—	—	—	—	—
O	+ 85	+ 22	− 23	+ 84	+ 28	+ 28
P	+ 84	+ 21	− 30	+ 19	+ 51	− 18
—	—	—	—	—	—	—
—	—	—	—	—	—	—

M = Firmness, N = Springiness } Assessed by pressing the thumb into the outer surface of the cheese.

O = Firmness, P = Springiness } Assessed by pressing the thumb into a sample taken from within the cheese.

In the studies on Cheshire cheeses firmness and springiness emerge effectively indistinguishable in spite of the fact that these two forms of

assessment are not perfectly correlated. In the Cheddar cheese studie the qualitative distinction between firmness and springiness is we substantiated.

Some Critical Comments

An effective appraisal of factor analysis as a technique for examinin a set of complex relationships in data on foodstuffs can be carried ou under two main headings. The first concerns specifically what ha been achieved in these particular studies of cheeses. The second con cerns more general questions relating to the application of this tech nique to other food products. Factor analysis has a limited range o uses. There may be certain tasks which can best be accomplished b this technique. Perhaps there are others which can be dealt with a effectively, or even more so, by means of some more widely acceptabl technique.

It has been argued that the two investigations of Cheshire cheese have revealed an excellent degree of general agreement in successiv years. It has been suggested that the evidence points to the necessit of employing three dimensions in order to describe effectively the wa in which one Cheshire cheese differs from another. This refers only t broadly quantitative differences along lines which must be presume to be common to all Cheshire cheeses. Idiosyncrasies do not emerg from this form of treatment, although they may exist and they may b important. Good reasons have been advanced for using a measure o surface hardness and the subjective assessment of crumbliness to repre sent the first two dimensions. The difficulties in the way of effectivel identifying the third dimension have been discussed, and in spite of wish to close this issue, the real nature of the third dimension must b left as unsettled. Such would seem to be the main conclusions whic have emerged from these time-consuming analyses. However, to hav established the multi-dimensional nature of the quality of a Cheshir or a Cheddar cheese is important.

Although no one form of analysis by any means exhausts the man facts lying hidden in the original data, it is not the function of th present article to examine what has been achieved by other method The question arises whether conclusions similar to those which hav emerged could have been drawn from less time-consuming statistica techniques. Perhaps it is true to say that all types of multivariat analysis are time-consuming, but the necessity for successive approxi mation by repeating the whole analysis—a regular part of the computa tional procedure in factor analysis—makes it especially so. Facto analysis could readily be adapted to the electronic computor, an much of the labour involved would thus be eliminated. As a techniqu for the broad classification of a set of empirical variables without fore knowledge of the nature (structure) of their interdependence, facto analysis seems to offer an approach which cannot be replaced by mor conventional statistical techniques. With the same original data

might have been possible to employ the mathematically more rigorous type of analysis into 'principal components' (see Hotelling[7]). In this case there is no need to attempt to give meaning to the principal axes which emerge. Since in the present studies emphasis has centred upon tests or assessments rather than factors, perhaps the facts which have emerged might just as well have been expressed in terms of this alternative and more rigorous approach. An analysis along such lines would be essential to test whether this statement is correct or not.

Basic assumptions enter into a technique at a number of different levels. Assumptions regarding the distribution of the original measures are implicit in the use of the product-moment correlation technique. Further steps in the analysis are based upon the assumption of a number of linear relationships, the effects of which can be added together in a simple manner. To what extent are these and other assumptions sustained by the data being analysed? Some writers believe, largely without proof, that departure from normality in the original measures and minor departures from linear relationships between pairs of variables are of little consequence in factor analysis. However, with this as with any other statistical techniques it is necessary to enquire from time to time whether the basic assumptions implied in the statistical models concerned are satisfied by the data. Unfortunately when a new field of enquiry is being opened up there are so many points at which assumptions have to be made that it is impossible to test whether more than a fraction of them are adequately satisfied. In the present series of investigations it was not possible to examine in detail the distributions of all the measures, nor was it possible to examine whether the relation between all pairs of variables departed significantly from a linear one. In one respect the subjective assessments certainly differed from the instrumental measurements. Since the subjective assessments represented an average of three persons' judgements it is important to note that there are differential restrictions on the possible variance at different parts of the rating scale. Extreme values could be achieved only by perfect consistency between raters, whereas values near the middle of the permitted range could be achieved in a variety of ways. It is impossible to say what influence this might have on the subsequent analysis, but there seemed to be no reason to distinguish between the instrumental tests and the judgements in the final analysis.

Apart from the measure of surface hardness, much remains to be achieved both with regard to the development of additional instrumental tests and with regard to the further standardisation of the methods of making and recording the essential subjective judgements. In view of the lack of finality of such developments, and in view of the many unarticulated assumptions which may be made, it seems reasonable to suggest that the simpler the statistical techniques employed, the better. It used to be fashionable to talk in terms of operational definitions and to say that a particular test itself largely defines what it

measures. However, in connection with both the subjective judgeme and the instrumental measurements which can be made upon chee it is important to stress that an operation is not an isolated event. Professor Meredith[9] has indicated operation, operator, and that whi is operated upon are inseparably related. Clearly, much remains be learned about these three aspects of the assessment of foodstuf Although some light may have been thrown by factor analysis upo such problems with special reference to cheeses, perhaps more in portant advances might be made by fashioning statistical techniqu around more clearly defined problems and around the actual (em pirically determined) properties of the data being accumulated.

Acknowledgements

I wish to thank Dr G. W. Scott Blair and Dr W. G. Wearmouth fo helpful discussion, although the opinions expressed are my own. Th editors of the *British Journal of Applied Physics* and of *Dairy Industrie* also the North Holland Publishing Company, Amsterdam, Holland have given their permission to reproduce information which ha originally appeared elsewhere.

REFERENCES

[1] Baker, G. A. (1954). 'Organoleptic ratings and analytical data for wines analyse into orthogonal factors', *Food Research*, **19**, 575.

[2] Cronbach, L. J. and Gleser, G. C. (1953). 'Assessing similarity between pr files', *Psychol. Bull.*, **50**, 456.

[3] Eysenck, H. J. (1952). 'Uses and abuses of factor analysis', *Applied Statistics*, 45.

[4] Harper, R. (1955). 'Fundamental problems in the subjective appraisal of foo stuffs', *ibid.* **4**, 145.

[5] Harper, R. and Baron, M. (1948). 'Factorial analysis of rheological measur ments on cheese', *Nature*, **162**, 821.

[6] *Idem* (1951). 'The application of factor analysis to tests on cheese', *Brit. J. App Phys.*, **2**, 35.

[7] Hotelling, H. (1933). 'An analysis of a complex of statistical variables in principal components', *J. Educ. Psychol.*, **24**, 417 and 498.

[8] Kendall, M. G. and Babington Smith, B. (1950). 'Factor analysis', *J. Statist. Soc.*, B, **12**, 60.

[9] Mekedith, G. P. (1952). 'Basic mental operations', unpublished address British Association, Belfast (Section J).

[10] Ostle, B. and Tischer, R. G. (1954). 'Statistical methods in food research *Advances in Food Research*, **5**, 161.

[11] Scott Blair, G. W. (Ed.) (1953). *Foodstuffs, their Plasticity, Fluidity and Co sistency*. North Holland Publishing Company, Amsterdam.

[12] Thurstone, L. L. (1947). *Multiple Factor Analysis*. University of Chicago Pres

[13] Vincent, D. F. (1953). 'The origin and development of factor analysis', *Appli Statistics*, **2**, 107.

[14] Wearmouth, W. G. and Baron, M. (1953). 'The relation between the mechan cal properties of Cheddar cheese and graders' judgements', *Dairy Industrie* **18**, 1094.

FIELD WORK IN SAMPLE SURVEYS OF HOUSEHOLD INCOME AND EXPENDITURE

DOROTHY E. COLE

Department of Applied Economics, University of Cambridge

In this article Mrs Cole discusses ways in which the efficiency of financial surveys might be improved, basing her recommendations largely upon experience gained in a recent survey of household income and expenditure.

Introduction

The last few years in this country have seen a considerable increase in the use of sample surveys for the collection of data on personal incomes, saving, and expenditure. There have recently been three major inquiries: the Ministry of Labour household budget inquiry; the Oxford Institute of Statistics Savings Surveys; and the project of the Department of Applied Economics, Cambridge, the Social Accounts of Cambridgeshire. (For a general description of the methods and objects of these three inquiries see references.[1, 2]) But while there has been considerable discussion about the design for such surveys, relatively little has been said about two other important aspects of the direct collection of income and expenditure information, namely the reliability of the material collected and the cost of collection. The existence of non-sampling error, while recognised, has not received very much attention, probably because of the difficulty of measuring it. In the course of the Cambridgeshire survey, however, it has been possible to form some views about the ways in which certain kinds of non-sampling error could be reduced, as well as to build up some data on costs.

I shall be concerned with three questions. Firstly, what can be done to increase response in surveys of human populations (particularly financial surveys), since incomplete response may always be a source of error. Secondly, what can be done to reduce recording errors arising from the mistakes of both the interviewer and the respondent. Thirdly, what, if any, will be the effect of the suggested steps upon the cost of field work for such financial surveys. The material drawn upon is, in the main, that obtained in a large-scale budget inquiry conducted by the Department of Applied Economics, Cambridge, as part of a research project, the Social Accounts of Cambridgeshire, designed to develop methods for the direct collection of national income and social accounting information. The field work was carried out during the period April 1953 to March 1954.

The Nature and Importance of Non-response

As far as response rate is concerned the general experience of t three surveys referred to was very similar—that somewhere betwe two-thirds and three-quarters of the units approached could be p suaded to co-operate. In the Cambridgeshire survey we were worki with a sample of 4 000 addresses confined to the administrative coun but spread over the whole year, so that there were some eighty address to be completed each week. Interviewers were instructed to see a adults resident at the selected addresses (these numbered some 9 000 They were collecting, by means of an interview, information about a expenditure in a past period—the period varying with the type expenditure—together with details of either income or savings over th past twelve months. The average length of interview was forty-fiv minutes. Complete co-operation was obtained at 73·5 per cent of a the addresses in the sample. The response rate in terms of household which was the important figure from our point of view because th household is the unit that is being used in analysis, was slightly highe 75·1 per cent, since some addresses were found to contain more tha one household, which made it possible for an incomplete address sti to yield a complete household. Of all the individuals who qualified fo interview 78 per cent co-operated. There was no tendency for respons to decline over the year.

In the Ministry of Labour survey 67 per cent of all the household approached provided complete expenditure information. The defin tion of household used was very similar to that used in the Cambridg shire survey and the interviewing was spread over a year. But th Ministry of Labour was asking people to complete records of the expenditure. The response rate is usually lower with this method tha with the interview method. The first Oxford National Survey Personal Incomes and Savings carried out in 1952 obtained a respons rate (reweighted)* of 75 per cent of all income units and the secon in 1953, of 66·1 per cent. In most cases only the head of the 'incom unit' had to be interviewed.

TABLE I

Proportion of all households co-operating

Stratum No.	Description	Percentage
1	City—RV up to £30	71·3
2	City—RV over £30	64·3
3	Rural—RV up to £13	85·0
4	Rural—RV over £13	71·6

RV = rateable value.

* Reweighted because the sampling scheme used variable sampling fractions in differe strata and the response rate varied between strata. Our reweighted household response ra was 76·1 per cent.

Both the Oxford[3, 4] and Cambridgeshire surveys revealed a greater readiness to co-operate among people living at low rateable-value addresses than among those at high rateable-value addresses, and among people living in rural areas than in the town. The variation in the Cambridgeshire survey is shown clearly in Table I.

It is possible that some of the difference can be attributed to the interviewers. On the whole, different interviewers were working in the rural areas and the city, and the best rural interviewers were responsible for a much higher proportion of the total interviewing in the two rural strata than were the best interviewers in the city strata. However, there is still evidence for stratum effect, because the best rural interviewers achieved a higher average response rate than the best town interviewers (see page 53).

It is perhaps worth noting that in Cambridgeshire experience does not lend support to any suggestion that the poorer response obtained in the high rateable-value strata arises from a lower response from households living more than one to an address. If anything, the following data suggest the contrary:

TABLE II

Percentage of households co-operating by number of households per address

Stratum No.	Number of Households per Address		
	1	2	3 or more
1	69·9	78·4	75·0
2	58·3	59·2	80·8
3	84·6	90·7	100·0
4	66·5	78·9	95·5

Although the level of response obtained in these surveys is encouraging, the position cannot be regarded as satisfactory. To take the information obtained as being representative of the whole population still involves the assumption that the spending or saving habits of those who refuse are not different from those who co-operate, even when we know that the people concerned are different in at least this one respect. It is, of course, possible to test for differences of known characteristics between the final sample and the whole population, as for instance Hill, Klein, and Straw have done for the 1953 Oxford Survey.[5] Some of these characteristics may in fact be closely associated with the variables that are being investigated, and the results of such tests may be very good guides to sources of error. But this is no substitute for efforts to increase the response rate.

Normally non-response in surveys of this kind is composed of two elements: addresses at which no one at all is seen or where a particular

individual is never contacted; and addresses at which people refuse co-operate. In the Cambridgeshire survey someone was seen at ev address visited (uninhabited addresses were abandoned only af neighbours, etc., had confirmed that no one was resident). So *individuals* were written off if repeated attempts to see them had fail (but even then only after consultations with the research staff). In the cases, however, it was felt that to draw a distinction between no contact and refusal would be extremely arbitrary. Non-availability often only an excuse to avoid giving a direct refusal. As a resu individuals who were not seen were treated as refusals, and the hous holds of which they were members then became refusing household

The elimination of non-contact at addresses can be attributed three main factors. The first was the result of working in a small are over a long period of time. If people were away temporarily when th first call was made, interviewing could be postponed until their retur This will introduce an element of bias into the estimates of expenditu outside the county which are being made, and for many surveys it clear that such a procedure would be impossible. An extension of fo weeks was needed to complete outstanding addresses at the end of th survey.

In most surveys quite a large amount of non-contact is due to peop being away in hospital ill, or being permanently sick at home, mentall deficient, etc. We were trying to interview people 'normally resident i.e. those who had lived or expected to live at the selected address f more than six months of the year. Thus someone ill in hospital wou either be away temporarily, in which case interviewing would be pos poned; or he would be away indefinitely and would not be regarde as normally resident at the selected address. (A separate samplin scheme was used for persons normally resident in institutions.) The must have been some temptation for interviewers to interpret 'indef nitely' rather flexibly in order to avoid having to postpone interviewin As for the people who were normally resident but were incapable being interviewed, our criterion was whether or not their financi affairs were handled by some other person in the household. Whe they were, information was obtained from that person. In some cas this might mean that we would obtain expenditure information b cause all purchases were made by another member of the household but that the income or savings information would be missing, since th was still the sole preserve of the sick individual. (Our criterion f response, it should be noted, was completion of the expenditure data In cases where no information was obtained the result was recorded a refusal.

Further, no limit was set upon the number of calls which an inte viewer was to make at an address in order to establish initial contac Since it is quite common practice with survey organisations to ma only three calls, the following table is interesting, because it sugges that the additional calls were important in establishing contact:

Call	*Percentage of all households contacted*
1st	62·0
2nd	22·3
3rd	9·0
4th	4·1
5th	1·6
6th + ...	1·0

Nearly 7 per cent of households were seen only after the third call. It should be pointed out that working in a relatively small area, with all our interviewers resident within it, probably made the cost of the extra calls very small.

By far the most important element in non-response is, however, the refusal to co-operate. There may well be something in the view that in financial surveys there is a 'hard core' of non-response, of people who refuse, on principle, to discuss their private affairs. But the variation in the response rates achieved by different interviewers in the Cambridgeshire survey would suggest that we have not yet reached such a hard core, and that much more attention needs to be paid to the role of interviewers. Among twenty interviewers working mainly in the city (where we have seen the response rate was lower) the proportion of *addresses* where complete co-operation was obtained varied from 51 to 85 per cent, with an average of 67 per cent. (It is easier for administrative reasons to discuss individual interviewers' response rates in terms of addresses.) In the rural areas the variation was not so great, but whereas the average co-operation was 81 per cent, the individual results of twelve interviewers ranged from 68 to 89 per cent.

The five best interviewers in the rural areas had between them an average response rate of 82·6 per cent compared with the 81 per cent achieved for the whole rural sample. In the town the five best had an average of 76·3 per cent compared with the overall average of 67 per cent. Had these ten interviewers, therefore, been able to work the whole sample for us, or had we recruited more carefully, we should have finished with a substantially higher response rate than we in fact did.

The Interviewer and Non-Response

Emphasis is here placed upon recruitment, for while it is true that certain techniques can be taught in training—for instance how to recognise and answer certain objections which are raised time and time again by people who are asked to co-operate—nevertheless these techniques, when acquired, have still to be applied in practice. For this it is clear that certain features of personality are essential which cannot be taught in a relatively short training period, although they may be developed after much self-discipline. The two most important of these features are what, for want of better names, we may call a sense of security in society, and emotional stamina.

The sense of security must apply to the interviewer's social positi generally and to the particular work of interviewing. To carry budget inquiries successfully the interviewer must have no misgivi about the propriety of asking for 'confidential' information. But m than that, the interviewer must be able to establish good relationshi with all kinds of people and to regard all people, whatever their incon their occupation or their social status, as equally important. Soc misfits in a class sense will have difficulty in doing this. For instan it is easy to be over-impressed by position or wealth, or shocked a condescending about poverty, if you aspire to a higher place in t social scale. An interviewer who feels compelled either to establish h superiority (intellectually, socially, financially) or to stake her claim equality of status cannot devote all her energies to understanding t outlook, and the possible misgivings and fears, of the person whom s has to persuade to co-operate. Self-consciousness arising from cla differences of any kind will introduce an element of artificiality into t interviewer-respondent relationship. It is perhaps significant that o of our most successful interviewers was Canadian.

We interviewed an applicant, middle class in origin, whose mo recent employment had been as a welfare officer. She appeared capab and pleasant. Once or twice during the course of the interview and t subsequent training she commented on how interesting the work w and how willing she would be to do the job voluntarily. In fact, on in the field, she had an extremely low response rate, and finally l suddenly and without notice. Two anecdotes which she related abo herself cast a good deal of light on the reasons for this. The first w a description of an interview with an academic figure—a scientist whom she was unable satisfactorily to explain our sampling metho Instead of recognising that she could not deal with his doubts a objections on his level, she allowed herself to be cross-questioned befo the man finally refused to co-operate. This and a somewhat simil experience made her angry and convinced that university people we difficult to interview. Needless to say, it was not uncommon for inte viewers to find themselves somewhat out of their depth with universi people who were genuinely interested in our methods. But the confide interviewers finding themselves in this situation had no misgivin about admitting their inability to provide detailed explanations, a referred people to the organisers for further information.

The second story of this same interviewer concerned an old-a pensioner living in a tumbledown cottage, who said she would he 'because it makes a job for people like you'. The interviewer's reactio to this was to think: 'Why, we'—meaning her husband and herself— 'could buy you up, lock, stock, and barrel and never notice it!' The was also the interviewer who characterised those who owned proper or appeared wealthy as the 'snob' type who exhausted her 'with the suspicious attitude', or yet another who appeared to consider anyo earning less than £10 a week as 'unintelligent'.

Unfortunately, it is precisely the people with 'social aspirations' who are often most attracted to interviewing, a 'genteel' occupation, particularly in Cambridge, where it meant being employed by the University. At the same time, it is not easy to detect social prejudices of this kind in general conversation or to stipulate any formal qualifications that will ensure the right approach on the part of the interviewer. But if one is aware of the problem a good deal can be learnt when first seeing applicants and also during any training period, by steering the discussion on to certain topics. The reaction of the potential interviewers to the question whether or not response is likely to vary with income, or why people are likely to be unwilling to co-operate, may prove a very useful guide to the presence or absence of this sense of security.

By emotional stamina, the second of our two factors, I mean the ability of interviewers to avoid becoming emotionally involved in the personal lives of those they are interviewing, and to stick at the job without being too personally affected by the results. A clear example of the way in which such emotional stamina works is in the case of repeated calls at an address. It is a well-known gambit for reluctant respondents to avoid taking a decision by being continually out or avoiding the interviewer. Repeated calls may be necessary in this situation to persuade respondents that the survey is serious and that the interviewer means business. All too often the poor interviewer becomes self-conscious about re-calls, feels that she is making a nuisance of herself, argues that further pressing will not be any use. There is, it is true, a real problem of knowing when to stop, to avoid being the kind of nuisance which endangers the public relations of the survey. But generally the fault of interviewers is not to press hard enough.

It must be said, however, that emotional stamina is even more difficult to detect in the early stages of selection than emotional security. Field work is the real test and the use of a large number of practice interviews a great advantage. Refusals are not *necessarily* the sign of bad interviewing, and interviewers sometimes need convincing of this at the beginning; but at least it is possible to learn a good deal from their reaction to these refusals.

These characteristics of good interviewers only became clear to us in the course of the Cambridgeshire survey. When we were recruiting we were not as selective as it was subsequently proved we should have been. We were worried about the possibility of getting enough people to do the work. We had a restricted area to draw upon, and although we could offer a year's employment there was little prospect of similar work being found in the county afterwards. Even so we by no means accepted everyone who responded to our initial advertisement in the local press. We had applications from 120 people, of whom 70 were rejected out of hand. We trained 36 after interview, but 12 dropped out during or at the end of training. The remainder, together with an already existing group of interviewers who had worked on our pilot

inquiries, made a total force of 51. Of these, 17 stayed with us t whole year, while others—through dismissals or resignations—had average working life of about six months. More care in selection wo have reduced this wastage, but the high proportion of married wor among interviewers is likely to mean that turnover will always something of a problem.

The Nature and Importance of Recording Errors

There are four factors which may influence what we call recordi error in financial surveys—that is the extent to which the answ recorded and used in analysis deviates from the 'true' answer. Th are the method used in the survey (questionnaire or record keepin the design of the questionnaire or record book, and the human elemen the interviewer and the respondent. The first has already been brie discussed in relation to the Cambridgeshire survey,[6] and the second a complete subject in itself and will only be touched on here. In t main I shall be concerned with the last two, the role of the interview and of the respondent.

Very little is known about the magnitude (even relative) of erro from these sources. By their very nature some are not even detectab and most survey material is used on the assumption that such errors exist are randomly distributed. Some idea of the nature and frequen of the mistakes made by interviewers can be gained from a study of t results of an examination which our interviewers took about half-w through the survey, where each one was required to interview a me ber of the research staff armed with a set script. The artificial natu of the situation must, however, be stressed. For instance, the int viewer's use of prompts listed on a questionnaire is an importa matter for investigation; but with a set script it is very much a mat of personal judgement on the part of the examiner whether or not t prompting is considered sufficient to elicit the answer. Further, t imaginary situation used for the examination brought together in o interview most of the difficulties and complications that could be fo seen, whereas in real life it is extremely unlikely that they would occur together. If we discount this, four main types of error wo apparent:

(i) Misuse of the questionnaire—either through failure to ask all t questions recorded or to use all the prompts listed.

(ii) Failure to ask necessary supplementary questions to establi the definitions of such concepts as 'household' or 'normally reside

(iii) Failure to pick up inconsistencies and duplications on t individual form or between the forms of different members of t same household.

(iv) Inaccuracy in the recording of actual amounts of expenditu dates, periods, etc.

This list can be compared with the results of what must have been a very similar test of interviewers carried out by the Social Survey.[7]

Some indication of the quantitative effect of the misuse of the questionnaire and actual inaccuracies in recording can be given from our examination. Details were collected of the amount of the last electricity and gas bill paid by the respondent, but the form specifically stated that hire-purchase payments (for appliances bought from the electricity or gas boards) were to be excluded at this point and recorded in a hire-purchase section. Twelve out of twenty-nine interviewers ignored this instruction, with the result that the expenditure on electricity was overestimated. At one point the interviewers were collecting information about housing expenditure (rent, if any; rates; mortgage payments; and house insurance) in such a way as to enable a year's expenditure to be calculated. Only nineteen recorded the correct information. On the grand total of annual housing expenditure for twenty-nine cases (which should have been £2 479 10s.) the answer was only 16s. out, however, because overestimates (£19 9s. in all) almost exactly balanced underestimates (£20 5s.). Experience with the total of last week's food expenditure was similar. On the grand total for all twenty-nine cases of £121 3s. 11d. there was only a deficit of £2 2s. But not one single interviewer got nearer than sixpence to the correct answer, because of inadequate use of prompts, failure to apply certain definitions, and also (it must be said) poor mental arithmetic.

The application of definitions was notably weak throughout the examination. No fewer than sixteen interviewers recorded the wrong number of persons resident at the imaginary address, or had mistakes in their household composition because of their failure to probe adequately. This was not necessarily a very reliable picture of what actually happens in the field on this particular point, because some, at least, of these mistakes would have been corrected as the interviewer moved on to see other households.

As a result of this test some interviewers were dismissed and the others were given further courses of training. It is clear that in training a great deal more emphasis needs to be placed on concrete illustrations of the kind of pitfalls involved in using the questionnaire, and of ways to avoid them. Testing itself is extremely useful in making interviewers aware of their faults. Even so, our experience suggests that not everyone is capable of benefiting from such training.

The Interviewer and Recording Errors

Nearly all our interviewers were housewives, most of them in the age group 35–50, with children of school age. We probably had more housewives than most survey organisations expect to employ because little mobility was demanded of the interviewers and little security was offered to make the job attractive to people who were absolutely dependent upon their earnings. There were only four male interviewers.

Almost all the interviewers had had jobs before—clerical and se
tarial work, selling, the civil service and armed services, teach
In many cases these jobs probably did not require any high degre
specialised training. As far as the housewives were concerned, sc
time had usually elapsed since their last jobs.

In general, the interviewers with some kind of previous train
involving real mental discipline were the ones who shone. They
cluded two women who had trained as social workers (although th
had never actually been professionally employed), a woman who h
worked for a firm noted for its rigorous office training, a school att
dance officer, and a teacher with a university background.

These good interviewers were distinguished by:

(i) Their acceptance of the necessity for a systematic way of gatheri
information;

(ii) Their ability to follow directions;

(iii) Their grasp of the complicated details of the questionnai
form.

The failure of the others was due, to a greater or less degree, to
impatience with a formal system of collecting facts. This made
difficult, if not impossible, for them to master the questionnaire. Hous
wives unfamiliar with work that requires mental descipline, who f
years had been making decisions on commonsense matters and taki
responsibility for them, were all too ready to form their own judgemer
of what was important to the survey (particularly when it was a surv
about household expenditure) without taking the trouble to make su
what information was really being sought. Not only does the int
viewer with the trained mind easily take the initial step of acceptir
the questionnaire and becoming familiar with its details, but she al
recognises the limitations of her own understanding, and when fac
with an unusual or complicated situation leaves the way open f
return after consultation with those in charge. The presence or absen
of the trained mind lends itself more easily to formal testing than th
other interviewer qualities we have discussed so far. Clerical tes
which involve the processing of material, the carrying out of instru
tions, the completion of simple sums, etc., can be used at the earli
stage for weeding out applicants. Further tests with the actual questio
naire can be used during the course of training.

The Respondent and Recording Errors

We have now to consider the respondent's role. Having once agre
to co-operate, he or she can be responsible for inaccuracies in the fin
answers because of a deliberate attempt to mislead, because of inabili
to supply the information required, or because of inadvertent concea
ment.

Deliberate concealment is difficult to deal with in any surve
The choice of the general method for an expenditure survey may

important. Quite a strong case can be made for the personal interview method on the grounds that since the respondent does not know the exact type and form of questions to be asked his replies are likely to be more spontaneous. There is less time for him to think up 'phoney' answers. Having learned something of the outlook of the person with whom she is dealing, the good interviewer can also sense the questions which are likely to be difficult and 'touchy' in the particular situation and take steps to allay any suspicions. On the other hand, it can be argued that record-keeping, because of its very impersonality, encourages people to be honest. It is probably easier, however, to be deliberately dishonest with a piece of paper than with another person; it is certainly easier to be consistently dishonest. Check questions with the interview method may reveal deliberate misinformation, whereas the record-keeper who is more aware of how the information he is providing should tie together may cover his traces more effectively. But in the last analysis, if people want to lie they will do so.

The distinction between deliberate lying and 'inadvertent concealment' is in many cases a fine one. It is a commonplace for expenditure surveys to produce underestimates of consumers' purchases of alcoholic drink and tobacco. From our own survey it would appear that Cambridgeshire households' expenditure on these two items is for alcoholic drink more than 50 per cent, and for tobacco 30 per cent below what we estimate to be the national average. (The national per capita figures of expenditure on alcoholic drink and tobacco have been multiplied by 2·9, the number of persons per household found in our survey.) Now while there is no reason to suppose that Cambridgeshire households will be average there is also no reason to suppose that they will be so very different. We can be fairly sure that a good deal of this underestimate results from people's unwillingness to reveal just how much they do spend on these two items. But there will be some people who do not admit even to *themselves* how much they spend. It is all too easy to pretend to oneself that one buys one packet of cigarettes a day 'on average' and to ignore the rather frequently recurring days when the one packet becomes two. Here the skill of the interviewer can help by the imaginative use of prompts—but if the illusion is important for the respondent, resistance to the interviewer's efforts may triumph. On other items where self-deception is not important, careful design of the form and its efficient use by interviewers can go a long way to overcoming the problems of 'memory'. There still remains much work to be done on deciding the most efficient list of prompts for helping people to recall their expenditure.

Finally there is the problem of the respondent who does not know the answers to the questions; sometimes he should be described as the respondent who can only discover the answer with a good deal of trouble. For instance, people with large incomes received from a number of sources who leave their affairs in the hands of solicitors or accountants may have very little idea of even their total income, let

alone the component parts. The breakdown between business an personal expenditure for sole traders is another example of the sa difficulty. An accountant preparing the business accounts may alloc such items as rent, fuel, etc., between business and personal exper ture. The respondent may first of all not be aware that such an allo tion is in fact made, or, if he is aware of it, he may still have little id of *how* it is made without reference to his business accounts. In all th cases the questionnaire and the interviewer can both help: the one asking questions which make clear what is being got at; the other convincing the respondent of the importance of accurate answers.

The Cost of Field Work

The cost of field work in any inquiry of this kind is considerabl several calls are often necessary to establish contact at any address, an the actual completion of the questionnaire is itself a lengthy busines It is difficult to say how the cost of the Cambridgeshire survey woul compare with a national survey. The size of the sample was ver similar, being based on expected variances which for items of housekol income and expenditure are likely to be of comparable size for a sma area as for the whole country. Travelling and subsistence, on the oth hand, are likely to have been lower, although not necessarily ver much lower than for a national body with a regional organisation.

Total field work cost just over £5 000, that is, 25s. per selecte address in the sample, and 33s. per co-operating household. This figu relates only to the direct interviewers' costs and includes no allowan for salaries of research staff or for general administrative control. O this, total training costs (including expenditure on those who did n complete the course) were £240, or less than 5 per cent of the tota Interviewers' wages at 3s. 6d. per hour amounted to £3 200, leavi just over £1 500 as the cost of travelling (including mileage allowan for cars), subsistence, and miscellaneous expenses.

It can be seen that training was a negligible part of total cost. complete training of eighteen hours and two practice field interview cost £6 18s. per person. To judge from our experience, longer trainin and more preliminary work in the field are necessary to improv recording accuracy. This might double the training cost, althoug greater selectivity of applicants would in part offset this. In any ca the final result would not be appreciable compared with the total co of field work and would appear well worth while.

The main burden of the argument has, however, been concerne more with the importance of selecting for interviewing work peop with the necessary initial qualities rather than with the kind of trainin they should receive. Here one is faced with the question whether it possible to find suitable people. Rates of pay may well have somethin to do with this, since interviewing has been regarded (and paid) as relatively unskilled occupation. Many survey organisations pay lower rate than can be earned by competent shorthand typists, an

lthough the technical qualifications needed in an interviewer are less, he intellectual qualities required may be much higher. On the other and, an improvement in the status of the interviewer in the survey rganisation and her involvement in other aspects of the survey might nake the work more attractive to the right type of people.

A major increase in interviewers' pay without any other changes vould clearly affect field costs considerably. Our £3 200 represented ome 18 000 interviewing hours. Payment at 4s. 6d. an hour would ave increased our total field costs by 18 per cent. But it is unlikely hat cost per completed household would be increased by anything like he same proportion. It is true that a higher response rate will mean nore time to complete interviewing at some addresses. In so far as ood interviewers work hard and consistently, however, this cuts costs y reducing travelling time. Good interviewers are also likely to take ewer calls to establish contact and get an initial response.

Conclusion

The experience of the Cambridgeshire survey suggests that the time as come for much greater attention to be paid to the reduction of on-sampling error in financial surveys. It has suggested that much reater care needs to be taken over the selection and training of inter-iewers for this kind of work, and has also indicated certain attributes vhich the good interviewer must possess. Further discussion of these entative conclusions in the light of the experience of other survey rganisations would prove invaluable at this stage.

Acknowledgement

I am much indebted to a report on interviewers prepared by Mrs Elizabeth Johnson, who helped us with the Cambridgeshire survey oth as an interviewer and as an organiser of the field staff.

REFERENCES

Lydall, H. F. (1954). 'The methods of the savings survey', *Bull. Oxf. Univ. Inst. Statist.*, **16**, Nos. 7 and 8, or *British Incomes and Savings*, ch. 7, p. 194, Blackwell, Oxford.

Utting, J. E. G. (1954). 'Sample surveys for household income and expenditure information', Schweizerische Zeitschrift für Volkswirtschaft und Statistik, **90**, No. 3.

Lydall, H. F. (1952). 'National survey of personal incomes and savings', *Bull. Oxf. Univ. Inst. Statist.*, **14**, Nos. 11 and 12, 372, or *British Incomes and Savings*, p. 226.

Hill, T. P., Klein, L. R., and Straw, K. H. (1955). 'The Savings Survey 1953: response rates and reliability of data', *Bull. Oxf. Univ. Inst. Statist.*, **17**, No. 1, 92.

Idem. (1955). *op. cit.*, 109.

Utting, J. E. G., and Cole, Dorothy (1953). 'Sample surveys for the social accounts of the household sector', *Bull. Oxf. Univ. Inst. Statist.*, **15**, No. 1, 6.

Fothergill, J. E., and Willcock, H. D. (1955). 'Interviewers and interviewing', *Incorporated Statistician*, **5**, No. 4 (Supp.), 44.

NOTES AND COMMENTS

Readers may be interested to hear of recent quality control developm in Holland. A Quality Control Service for Industry has been institu Stichting Kwaliteitsdienst voor de Industrie (KDI), 24 Wassenaarsev The Hague. The foundation works in close co-operation with other or nisations; it was financed in the first place by 'conditional aid' funds hopes to become self-supporting in a year or so. Its avowed purposes are dissemination of information, promoting the introduction of quality p grammes, and training in the techniques of quality control.

In co-operation with the Netherlands Statistical Society the KDI is pu lishing a bi-monthly journal *Sigma* with a circulation of 2 000. Four numb had been issued by the end of 1955. Some meetings and training cour have already been held. These were arranged by some of the co-operati organisations, by the Research Associations, or by individual firms. KDI co-operated in these and now intends to provide standardised training some hundreds of quality control operators. These courses will be decentr ised, will be by specially trained instructors, and each will deal with o one industry or with one form of production. The more advanced traini including that of the quality control analyst, will be left to the universi and to the Netherlands Statistical Society, which organises appropri courses at the Institute of Statistics.

★ ★ ★

Dr W. J. Youden, besides being the author of a very acceptable int ductory book on statistics for chemists, writes a 'column' that appe regularly, at two-monthly intervals, in *Industrial and Engineering Chemi* under the general heading of Statistical Design. These short articles addressed to industrial chemists and managers, and judging from those t have appeared so far, are aimed at increasing awareness of what statis can do to make experiments more effective. Such penetration of statis into the technical literature of the fields of application should do much the extension of applied statistics.

★ ★ ★

A one-day conference on 'Statistical Methods and Industrial Experimen organised by the Birmingham Group of the Industrial Applications Sect of the Royal Statistical Society, is to be held at the Birmingham College Technology on Wednesday, 9th May 1956. The object is to promote inte among scientists and technologists in the use of statistical methods in exp mental work; emphasis will therefore be on general principles and illustra examples rather than on the finer details of technique. Dr G. E. P. Box Dr H. C. Hamaker have agreed to give lectures and other speakers are be arranged. Enquiries should be addressed to the Hon. Sec. of the mingham Group, W. N. Jessop, c/o Operational Research Departm Courtaulds Ltd, 166a Lockhurst Lane, Coventry.

QUESTIONS AND ANSWERS

Readers are invited to submit questions or to deal shortly with points of general interest by supplying both the question and the answer.

uestion

When observations are confined to the integers 0, 1, 2, 3, . . . a transfor-ation of the type log $(x + c)$ is often recommended, where x is an observa-on and c is some positive constant. How should c be chosen?

nswer

Such transformations are employed in order to permit the application of nalysis of variance techniques, and we should then like the transformed bservations to be linearly dependent on the various factors involved in the xperiment and to have a constant error variance; the first requirement is *dditivity* and the second *homogeneity*. Since the observations and their trans-rms have discrete distributions, neither can be exactly normal. However, ne F-test is fairly insensitive to non-normality, so we can concentrate on ecuring additivity and homogeneity. In general we cannot obtain both by single transformation, but we hope that there exists a value of c such that oth are satisfied approximately. If no such value exists, the problem equires expert attention.

First of all, consider what happens to the transformation log $(x + c)$ hen c is small and when it is large. When c is much smaller than 1, g $(x + c)$ is large and negative for $x = 0$, relatively small and positive r $x \geqslant 1$; changing the origin and scale, the transformation is equivalent replacing all the values of $x \geqslant 1$ by 1, leaving the zeros unaltered. When is large compared with the values of x:

$$\log (x + 1 + c) - \log (x + c) \simeq 1/c$$

lthough the origin and scale are changed, the transformation is equivalent leaving all the values of x unaltered. These two points were made by ukey and Moore (*Biometrics*, **10**, 562) and indicate that we can confine ur attention to intermediate values of c.

Consider next the requirements of additivity. Tukey (*Biometrics*, **5**, 232) as proposed a test for additivity in which one degree of freedom is isolated om the residual sum of squares and compared with the remainder; and in s article with Moore (*loc. cit.*) it is suggested that the test should be made r different values of c, thus determining the range of values in which non-gnificant results are obtained. In this range any transformation will tisfy the first requirement within reasonable limits.

As regards homogeneity, suppose that we have sets of observations repeated nder the same experimental conditions and can thus examine the relation etween mean and standard deviation of a sample. For observations con-sting of counts these quantities usually increase together, and the object

is to find c such that the standard deviation of the transformed observati is independent of the mean or nearly so. The following method is d.e Kleczowski (*Ann. Appl. Biol.*, **36**, 139). For each of k samples comput mean m and standard deviation s. Calculate

$$\bar{m} = \Sigma m/k; \quad \bar{s} = \Sigma s/k; \quad M = \Sigma m^2 - (\Sigma m)^2/k; \quad T = \Sigma ms - (\Sigma m)(\Sigma s)/k$$

Then $c = M\bar{s}/T - \bar{m}$.

R. L. PLACK

Question

In a factorial experiment involving two factors A, B and their interact A $\times$ B, the interaction is found to be not significantly larger than the exp mental error. Should it be pooled to provide a better estimate of this err

Answer

The fact that an interaction is not significant does not mean that it d not exist, only that it was not large enough to be picked up by the anal of variance. Unless the interaction term is actually zero, so that the fact A and B are additive, a procedure of pooling in all experiments where interaction mean square is non-significant will therefore lead to a bia estimate of error. Thus the interaction mean square should not be poo with the residual whatever the result of a test of significance. For a discuss of the bias due to pooling see T. A. Bancroft (*Ann. Math. Statist.*, **15**, 190)

R. L. PLACK

INDUSTRIAL APPLICATIONS SECTION OF THE ROYAL STATISTICAL SOCIETY

List and Reports of Meetings

BIRMINGHAM AND DISTRICT GROUP

5th October 1955. The Uses of Factor Analysis in Industrial Physics: G. W. Scott Blair

26th October 1955. A Method of Labour Turnover Analysis: K. F. Lane and J. E. Andrew

30th November 1955. The Role of Electronic Computers in Statistics: K. D. Tocher

GLASGOW GROUP

19th October 1955. Personnel Selection: P. A. D. Gardner

24th November 1955. The Use of Statistics in the Study of Market Developments: J. V. Gregg

SYNOPSIS: Some techniques for interpreting change in sales level. Differences in treatment of long-term and short-term market developments.

LEICESTER GROUP

20th October 1955. The Importance of the Elementary: B. P. Dudding

1st December 1955. Statistics and the Accountant: W. F. Harris

LONDON GROUP

7th October 1955. Quality Control in the Paper Industry: I. F. Hendry

Mr Hendry dealt with the problems connected with the measurement of a variable continuous in two directions, such as paper substance, which can vary along and across the web. He described the difficulties of applying a quality control scheme to such a system.

4th November 1955. Clerical Operations as Statistical Processes: R. A. Fairthorne

An immense amount of labour is utilised in the transference of marks from one sheet of paper to another, as the lecturer humorously described 'copy typing'. He developed a theory for these routines of observation, identification, and manipulation of statistical assemblies of material-marked objects, and showed its formal equivalence to translation of number languages. Mr Fairthorne finally described a number of points which must be observed when creating library classification systems.

2nd December 1955. Linear Programming in Practice: E. Kay and W. E. Duckworth

The speakers dealt with the problem of optimum usage of the scrap from a series of casting operations to make up fresh batches of slightly differing alloys. It was shown that this method is slightly more efficient than a rule of thumb method in terms of metal used, but makes a marked reduction in the hold-up time of metal and in the time spent in calculating prescriptions.

MERSEYSIDE GROUP

4th October 1955. The Applications of Electronic Computing in Industry: B. B. Swann

1st November 1955. Labour Wastage in Industry: H. Silcock

6th December 1955. Monte Carlo Methods and Industrial Problems: W. N. Jessop

NORTH-EASTERN GROUP

12th October 1955. Snap Observation Techniques: K. T. Boyd

There are many objections to classical time study with a stop-watch. Mr Boyd showed how snap observations on a randomised programme or inspection can provide a satisfactory alternative. The ratio-delay, or proportion of time a machine is not working, can be calculated with specified accuracy, or estimated from a nomogram relating the number of observations and number of 'machine stopped' observations to the ratio-delay and its 3σ limits. The method can be extended to study the frequency of the various causes of stoppage.

9th November 1955. Some Applications of Statistics to the Improvement of Colliery Operations: B. H. P. Rivett

The Field Investigation Group of the National Coal Board has tackled a number of problems requiring statistical methods. Mr Rivett first described a method of comparing underground telephone layouts as warning systems. Monte Carlo methods and the ACE were used to produce sigmoid curves of the proportion of men warned by a given time, from which the more effective warning systems could be identified. Other topics discussed were the design of an experiment in incomplete blocks to compare belting materials, and methods of improving the speed of tunnelling in coal-mines.

SHEFFIELD GROUP

29th September 1955. Ultimate Capacity of a Slabbing Mill: H. G. Jones

27th October 1955. Statistics without Computers: H. R. Watkins

24th November 1955. Consumer Preference Testing: N. J. Squirrell

SOUTH WALES GROUP

6th October 1955. Opinion Survey: Mark A. Abrams

Speaking on some of the theories behind opinion surveys, Dr Abrams contended that many of the basic assumptions were false. He showed by example that man will change his opinions depending on the role he is playing at any time; that, far from being rational, he has many conflicting attitudes and that most people have strong opinions about a particular subject without knowing any facts upon which to base them. Opinion surveys also assume that the average man is literate whereas, in fact, most people manage from day to day on 600 words; and when asked their opinion they often give meaningless answers because of their poor vocabulary. Dr Abrams suggested that opinions must be defined in terms of behaviour, which itself is a rationalisation of one's private motives. He went on to describe recent methods of motivation research which had been found to give better information than a direct questionnaire.

23rd November 1955. Some Facts concerning Road Accidents in Great Britain: F. Garwood.

The address was substantially that given to the Study Section on 13th February 1952 and reported in *Applied Statistics*, **1**, 145.

STUDY SECTION OF THE ROYAL STATISTICAL SOCIETY

List and Reports of Meetings

LONDON GROUP

12th October 1955. A Research Design for studying the Effects of Television: W. A. Belson

9th November 1955. The Scope and Limitations of Econometrics: L. Klein

It is hoped that these lectures will be published in later issues of *Applied Statistics*.

BRISTOL GROUP

20th October 1955. Personnel Statistics: H. E. Bishop

Mr Bishop discussed the influence of the prospects of promotion upon recruitment and retirement policy in an organisation, and also considered the effect of labour wastage and unstable age distribution on the long-term personnel planning.

17th November 1955. The Application of Statistical Methods to Market Research Data: E. S. Holdsworth

Mr Holdsworth showed how the finished product of a market research survey is used by a particular industry, and how checks can be applied by the industry to assess the amount and type of bias involved in the survey. He also indicated the type of information required, in the form of age and class distributions of preference for brands or types of product, and the methods of presentation to facilitate digestion by the executives.

BOOK REVIEWS

Lancelot Hogben: *Chance and Choice by Cardpack and Chessboard.* Volume 2. London: M
Parrish, 1955. Pp. 491. 70s.

Professor Hogben has now completed his enterprising venture in the fiel
of statistical education. His object has been to clarify the basic assumption
and expose the theoretical structure of many techniques important in applie
statistics, without exceeding the level of mathematical knowledge repre
sented by the Advanced papers of the General Certificate of Education. I
this volume he is largely concerned with analysis of variance, regression, an
goodness of fit; but he also offers an introduction to methods not so widel
accessible, such as factor analysis and sequential tests. The pattern of hi
approach is the same as in Volume 1. He begins with situations where al
the probabilities can be enumerated, illustrating them by coloured diagram
of cardpacks and other gaming devices; and he then proceeds to more diffi
cult problems involving continuous variation.

His account of the analysis of variance can be taken as a fair sample of th
advantages and shortcomings of his presentation. He distinguishes betwee
random and systematic models, examines the partition of sums of squares
and derives distribution theory for the appropriate significance tests; thes
operations take several chapters to complete, and perseverance is a necessar
adjunct to survival. The limitations of the technique as an instrument c
biological research are brought out by a devastating section on cuckoos
their eggs, and the non-existence of their nests; and he emphasises that th
principles were developed for small adjustments due to errors of observation
where unknowns are linearly connected, and that similar assumptions i
other problems should be made with caution. He does not discuss the powe
of the F-test to detect departures from the null hypothesis, its robustnes
under non-normality, or its sensitivity to differing variances. Such matter
are undoubtedly part of the rationale of the test, which is what he claims t
explain, and their omission may possibly be due to the mathematical restric
tions he has imposed; but as far as the question of power is concerned, in
ternal evidence suggests that he was simply unaware of its existence unti
most of the book had gone to press.

What he does discuss, however, is carefully argued with a wealth of vivi
terminology and numerical illustration; he always writes well, and his indi
vidual outlook is continually refreshing. There are plenty of exercises
although they peter out in later chapters; and a rather short index cover
both volumes.

R. L. PLACKET

Sir Ronald A. Fisher: *Statistical Methods for Research Workers* (Biological Monographs an
Manuals No. V). Edinburgh: Oliver and Boyd, 1954. Pp. xiv + 356. 16s.

The latest edition of this famous book, first published in 1925, differs littl
from its immediate predecessors, and, in spite of various changes and addi
tions, not a great deal from the early editions. Since statistical method ha
developed considerably over this period, and the composition of the bod

of its practitioners has changed, in a review of the book the questions naturally arise: how far has its value become historical rather than practical, and what should be its role in the present-day study of the subject?

Of its historical value there is no doubt: it was the expression of the impact of a highly original mind on a newly developing science, and the first (and in many ways still unsurpassed) connected exposition of that theory of small samples, pioneered by the author and 'Student', which made an honest woman of statistics. It is an essential part of the cultural heritage of all statisticians.

From the practical point of view there is no doubt that the early chapters on distributions, goodness of fit, and significance of means and regression coefficients are as fresh and relevant today as ever. The bulk of their contents forms an indispensable part of the equipment of every practising statistician, and I doubt whether these topics have been treated since in a more enlightening (to a mature mind) or elegant way. When this book was written, Karl Pearson's work had given the correlation coefficient a much more central position in the body of statistical method than it holds today; hence in the middle chapters it is given prominence excessive by modern standards. Generations of statisticians have found the chapter devoted to it a beautiful and stimulating piece of work, but have had no subsequent practical need to make use of the techniques involving z to which it is so largely devoted. Also the introduction to analysis of variance via intra-class correlation seems quaint and strangely circuitous to modern readers; yet the interpretations of analysis of variance which find favour now are given too, and the concepts of analysis of variance and degrees of freedom permeate the book and weld it into a unified whole.

The final chapter on statistical estimation is a concise communication and illustration of sophisticated but important ideas which not only are basic to the modern mathematical theory of statistics, but also have tremendous practical importance in those fields of research where much work goes into the production of even a single observation.

Statistical Methods for Research Workers is very concisely written, and although few mathematical formulae are to be seen, firm indications are always given of the theoretical bases of the techniques proposed. In spite of its shortness, a statistician who fully understood its contents would be well equipped even by modern standards. Yet this conciseness is a serious disadvantage to students or research workers making their first acquaintance with statistics, many of whom do in fact turn to this book for guidance. For example, they often tend to skip a reference which indicates an important idea or contains the essential germ of a technique as merely a somewhat obscure remark, and may be astonished later when it is pointed out to them as the answer to some problem which they thought was new. Comprehension, too, is sometimes baffled by the obscurity of some of the examples; allelomorphs, linkage, differential viability, coupling in repulsion, are unexplained terms outside the vocabulary of all but certain specialised biological research workers. These faults, serious to the self-instructor, matter much less when the book is used as an accompaniment to lectures; an extremely useful introductory course for modern research workers could consist of a series of lectures in explanation (with less emphasis on certain topics) and supplementary illustration of nothing more than this book. It could then be re-read (which is easy to do, and it does not lose interest on re-reading)

from time to time by the worker to refresh his memory and impress on him with ever-increasing clarity the modern statistician's way of looking at data.

G. H. JOWETT

M. S. Bartlett: *An Introduction to Stochastic Processes*. Cambridge University Press, 1955. Pp. xiv + 312. 35s.

Classical statistics is essentially a study of the stochastic aspects of static situations, with order in time or space largely irrelevant. This new branch of statistics, on the other hand, deals with situations which are changing in time or space, where the nature of the picture at any instant depends, for example, on what has gone before. Over recent years it has been developed in an *ad hoc* way by physicists studying statistical mechanics and the motion of fluids in turbulence; by mathematicians in games of chance; by electrical engineers in the communication of signals, the study of noise, and so on; and by statisticians studying population growth with births and deaths, sequential sampling, queues and renewals, and the correlation of time series. Professor Bartlett's book meets an important need in presenting a general treatment of the subject, showing the essential unity of the mathematical concepts and methods which pervade all these applications.

The early chapters deal with the mathematical theory of random walks and additive processes, Markov and multiplicative chains, in both discrete and continuous time, and are followed by a chapter on applications dealing with queues, renewals, growth of populations subject to births and deaths, and the transmission of characters in genetics. Next follows an outline of the stochastic calculus, covering almost certain convergence, mean square differentiability, and stochastic differential equations, bringing the reader into the stiller air of stationary processes which remain invariant in underlying probabilistic structure with the passage of time. It is shown how a stationary series, non-periodic though it may be, can be characterised in terms of a spectrum of waves of different periods, occurring with different densities, by passing it through the prism of Fourier analysis; the study of this spectrum is central in this chapter and in the subsequent one on prediction and communication theory. The final chapters cover statistical inference for stochastic processes in general (i.e. estimation of parameters and their accuracy from observations of processes in operation) and the correlation analysis of time series, with the attendant problems of finding a descriptive model for a series, testing its goodness of fit, and finding the form of its spectrum; the author's own recent researches are particularly in evidence here.

The standard of mathematical attainment required in the reader is high, though by no means so high as by Professor Doob in his recent book; an honours degree in mathematics would be about right. This limits the value of the book to industrial statisticians, the more so because the applicable formulae do not stand out from the text, and it is rarely obvious without careful study of a mathematical kind what sort of questions are answered about the various processes, or what the answers are. Demonstrating the application of some of the formulae could form a useful object for many a future article in *Applied Statistics*, like that by Mr Ramm in a recent issue.

It seems likely that this book, which is elegantly written and nicely produced, will become the standard work on the subject for advanced students in British universities and will find considerable favour abroad. It will form a stepping-off point for much important and practically useful research.

G. H. JOWETT

Hugh H. Wolfenden: *Population Statistics and their Compilation* (Revised Edition). Published for the Society of Actuaries by The University of Chicago Press, 1954. Pp. xxii + 258. 56s. 6d.

There must be few business statisticians—particularly among those concerned with marketing problems—who from time to time are not called upon to analyse and interpret population statistics. And, similarly, there will not be many practising economists, especially those concerned with conditions of economic growth, who are not confronted with demographic considerations. For these certainly, the *subject* of Mr Wolfenden's book, a revised edition of one originally published in 1945, is of considerable interest.

The early chapters deal with the collection of population statistics by means of censuses and the registration of births, deaths and marriages. As a review of the problems that have to be tackled by those responsible for the preparation of vital statistics, these chapters offer essential background to anyone undertaking serious work in this field. There are certainly many pitfalls in the use of population data which a study of these chapters might well help to avoid.

However, the middle chapters, and the bulk of the book, comprise largely a detailed compilation of specialised interpolation formulae geared to the types of functions actuaries expect to meet. Undoubtedly these are of considerable interest in the construction of mortality and life tables; but while of value to the actuarial student and demographers they will probably prove of little interest to others.

In Chapter XIV Mr Wolfenden deals with the interesting and complex problem of the measurement of reproductivity. The reader is given a brief review of the various methods that have been used and draws attention to the defects of the Gross and Net Reproduction Rates, which had achieved considerable vogue among demographers. Little, so far, in the way of a simple index of reproductivity has been put in their place. Yet such a measure is needed; it may be true, as Mr Wolfenden points out, that such indices 'cannot under any circumstances replace a detailed analysis of the mortality, marriage and birth rates by ages and duration as a means of examining the trends in marriages, the number of children per marriage, and the ultimate size of the population'. But few, interested in obtaining a reasonably reliable guide as to current population trends, have sufficient time or resources to digest this data.

The prediction of future populations receives rather less attention than the non-specialist reader might have hoped. Certainly there is a general review of the various methods that have been adopted by various authorities, and reference is made to the admirable paper by H. F. Dorn (*J. Amer. Statist. Ass.*, **45**, 311) on 'Pitfalls in Population Forecasts and Projections'. Clearly Mr. Wolfenden is not very happy about the validity of long-term projections of the general population, and this probably explains the brief treatment that they receive in his book.

The last chapter of the book deals with sickness data. It would seem that the author's fears about the reliability of sickness data obtained by 'enumerators without any medical background' is contrary to experience in this country, where valuable data were collected in the series of Surveys of Sickness carried out by the Government Social Survey. This scanty chapter can hardly be considered adequate as a discussion of the methods by which sickness data might be collected; it is a pity, too, that no opportunity was

taken of discussing the possibility of applying sampling methods, summarised by Dr W. Edwards Deming in an Appendix, to the collection of sickness data.

The author is at pains in his conclusions to point out that his book is intended to describe methods of compilation and not to discuss their interpretation or 'other allied problems which are more particularly interesting to the general statistician'. In brief, therefore, it may be a useful *vade mecum* for the actuarial student, but it is doubtful whether—apart from the first few chapters—it is likely to be of great interest to others.

E. SHANKLEMAN

PUBLICATIONS RECEIVED

Mathematical Models of Human Behavior. Proceedings of a Symposium sponsored by Dunlap and Associates Inc. and the Commission on Accidental Trauma, Armed Forces Epidemiological Board. Stamford, Connecticut: Dunlap and Associates Inc., 1955. Pp. vii + 103.

The following papers are reported in this book: A Group Preference Experiment, M. M. Flood; Some Problems in Stochastic Learning Models with Three or More Responses, R. R. Bush; Mathematical Models applied to Accident Processes, H. H. Jacobs; Ψ Stability: A New Equilibrium Concept, R. D. Luce; Norms and Habits of Decision Making under Certainty, J. Marschak; The Utility of Wealth, H. Markowitz; Theory of Elementary Predictive Behavior: An Exercise in the Behavioral Interpretation of a Mathematical Model, W. K. Estes; On the Multidimensional Analysis of Monotonic Single Stimuli Data, C. H. Coombs and R. C. Kao; Probability Estimates and Gambling, M. E. Jarvik; An Attempt to Predict Gambling Decisions, W. Edwards; Concluding Remarks, P. Lazarsfeld.

Contributi all'Analisi di Alcuni Fenomeni trattati nella Relazione annuale del Governatore. Servigio Studi Economici della Banca d'Italia. Rome, 1954. Pp. 232.

EXAMPLES OF INTERVIEWER VARIABILITY TAKEN FROM TWO SAMPLE SURVEYS

PERCY G. GRAY
Government Social Survey

Most people who carry out sample surveys are aware of the fact that the results are liable to depend to some extent on the interviewers employed. It is not always possible to measure the extent of the variability between interviewers, although steps may be taken to minimise its effects. In this article Mr Gray describes some measurements of variability carried out in connection with two surveys in this country.

ntroduction

It is well known that different interviewers working under equivalent onditions with equivalent respondents sometimes obtain different esults. Few results of the measurement of this interviewer variability n actual surveys have been published in this country. In the succeed- ng sections I shall give some examples of items where interviewer ariability has been examined. These are taken from two surveys, one ealing with noise in blocks of flats, the other with morbidity.

Noise Survey

The first and more recent example is taken from a survey[1] which was oncerned with the reactions to noise of the tenants in three groups of ats where the floors provided different levels of sound insulation. An nusual feature of the interviewing in this inquiry was that an inter- iew was sought in each flat of all blocks chosen for study. Two factors avoured this inclusion of whole blocks instead of isolated flats. In the rst place it was possible for the mobile laboratory of the Building esearch Station to measure the insulation values of only a limited umber of floors on a limited number of sites. In addition we wanted o know not only the complaints from a particular flat but also about ne possible sources of noise in the flats surrounding it. The economical ay of securing both sets of information was clearly to consider each at in a block both as a source of complaint and as a source of noise. hus the aim was to interview the housewife in each flat of the block.

The original plans for the inquiry included only two groups of flats, hich were all in the London area, and this part of the inquiry was arried out in December 1952. Group I, with the highest sound insula- on, consisted of 380 flats all in one estate. Group II, with medium ound insulation, consisted of 560 flats on three estates. Later Group II was added to the inquiry in March 1953. This group had the

lowest insulation and consisted of 599 flats on two estates in Glasg

With the interviewing clustered in this way it was a simple matte allocate interpenetrating samples to the interviewers for each grou flats. Thus each of the original group of twenty interviewers selec for this inquiry was allocated a sample of nineteen Group I flats a twenty-eight Group II flats. Later, when it was decided to include Group III flats in the inquiry, it proved impossible to use this sa group of twenty interviewers owing to the other commitments of Social Survey. Of the twelve interviewers who dealt with the Group flats only two came from the original twenty. As events turned out had some cause to regret this, as will be seen later. However, I sh first give the results for the interviews obtained by the twenty int viewers who dealt with the Group I and II flats.

Normally the distance between interviews makes it difficult to a for more than two recalls at any address where the informant is not home at the first call. Here the proximity of addresses made it possi to ask for unlimited recalling during the three days allocated to int viewing on each estate. We also established temporary headquart near each of the estates in turn and were thus able to exercise cl supervision of the work. This concentration of the interviewing, a perhaps the closer supervision, probably account for the unusually hi proportion of successful interviews.

	Number	Percentage
Interview obtained	922	98·1
Reason for failure:		
Flat untenanted	5	0·5
Refusal	7	0·8
Housewife repeatedly out or temporarily away ..	6	0·6
All Group I and II flats	940	100·0

The measures of interviewer variability given in Tables I, II, and are based on these 922 interviews obtained by the twenty interview The analysis of variance took the form

Source of Variance	*d.f.*
Between groups of flats	1
Between interviewers	19
Group-interviewer interaction ..	19
Residual	882

We are not concerned here with the differences between the two grou of flats, and these will not be given. Except in a few cases which noted in the tables the group-interviewer interaction was not significa With these exceptions I shall consider only the between-intervie effect. Estimates of the ratio of the between-interviewer variance to within-interviewer variance have been made for a large number

ariables. This ratio, F, may be expected to be distributed approxi-nately as the variance ratio.

TABLE I

Noise survey—'factual' items—
ratios of between-interviewer variance to within-interviewer variance

Attribute or Variate Analysed	F
Number of persons in household	0·88
Age of housewife	0·98
Housewife a worker	1·28
Housewife lived in a flat previously	0·89
Housewife's normal bedtime 11 p.m. or later	0·66
Household has a television set	1·21
Household has a vacuum cleaner	0·65
Sitting room has a carpet	2·01**

** Ratio is significant at 1 per cent level.

Table I gives the value of this ratio for eight variables of what might be called a 'factual' nature, which could be expected to show little interviewer variability. For seven items the ratio does not attain significance at the 5 per cent level, but for the final item, whether or not the sitting room had a carpet, the value is significant. (No attempt has been made to transform the variables in the case of proportions. It was felt that the ratios of the variances for the untransformed variables would be more informative, although this means that all the conditions for the F-test are not strictly fulfilled.) The 'carpet' classification was made as the result of observation by the interviewer and not by questioning. The precoded classification supplied was as follows:

Carpet	1
Rugs/mats and lino	2
Rugs/mats only	3
Lino only	4
None	5

nd the instructions stated that only one number should be ringed, the resence of a carpet being sufficient to cause 1 to be ringed. Examina-ion of the results, interviewer by interviewer, shows that the removal f one interviewer's results would reduce the value of F below the 5 per ent level of significance. This interviewer recorded only 5 carpets for 7 flats as compared with an expected 37 per cent. She had, of course, n excess of codes 2 and 3. It is unfortunate that this analysis was made oo long after the event to permit of recalls being made. This inter-iewer's results for other questions were not abnormal. Thirty-six per

cent of her flats had vacuum cleaners as compared with thirty-five p cent for the whole sample, which immediately suggests that her fi were as well carpeted as the rest. Two further observations, which a required entry into the sitting room, showed no abnormality: t suggests that the observations were in fact made. Questioning of t interviewer, now married but then on the point of marriage, sugge that she may have been a little preoccupied with thoughts of fitt carpets and adopted a too rigorous definition of 'carpet'. It is n clear that more care should have been taken on our part in defini the different classes of floor covering.

Thus for eight items for which a reasonably precise definition cou be given, seven showed no significant interviewer variability. In t remaining case the definition was at fault.

TABLE II

Noise survey—questions dealing with complaints and noise—ratios of between-interviewer variance to within-interviewer variance

Question Number	Description of Question	*F*
1	Are there any things about this flat that you find unsatisfactory? 'Anything else?' prompted after each reply. Number of complaints..	3·39**
4	Do you think too much noise comes through the.....? (Proportion answering 'Yes')	
	Ceiling	2·71**
	Floor	2·26**
	Walls	1·53
17	Can you usually hear people when they are talking in the .....? (Proportion answering 'Yes')	
	Flat above	3·35**
	Flat below	†
	Flat next door	1·75*
24	Could you sum up your opinion by saying whether you find noise from inside the building: very disturbing, disturbing, a little disturbing, or not at all disturbing?	
	Very disturbing ..	1·77*
	Very disturbing or disturbing ..	2·17**

* Ratio is significant at 5 per cent level.

** Ratio is significant at 1 per cent level.

† Group-interviewer interaction significant.

When we turn to the questions dealing with complaints or noise n such precise definitions were possible and a different picture is pr sented. Table II shows the first, the last, and two intermediate que tions taken from the main section of the questionnaire. All of thes show significant interviewer variability.

Open questions similar to 1, which invite any number of complaint are very liable to be affected by the interviewer. The question ma

produce a flood of complaints which the interviewer may find difficult to cope with—she may then prove selective in what she records, possibly omitting the more 'obvious' answers. The main purpose of the question was to provide a count of the number of complaints other than about noise and so to measure the general tendency to complain about the flat. It may be noted that such a count of the number of complaints presents certain difficulties at the classification or coding stage in deciding what constitutes a fresh complaint and not merely a different aspect of the same one.

It might be thought that by the end of the noise questions there would be no variation between interviewers but it persisted in the final summarising question, 24. Combining 'very disturbing' with 'disturbing' to form a single category did not eliminate the variability.

TABLE III

Noise survey—further set of noise questions—ratios of between-interviewer variance to within-interviewer variance. The variable used is the proportion hearing or disturbed by a noise (see below)

Noise	Hearing Noise from Flat			Disturbed by Noise from Flat		
	Above	Below	Next Door	Above	Below	Next Door
Wireless or television	†	1·76*	0·57	1·77*	1·60*	†
Baby crying	1·13	†	1·29	1·85*	1·48	1·06
Child's voice ..	1·97**	2·59**	1·59	1·67*	1·09	0·74
Grown-up's voice ..	1·26	1·61*	1·15	1·53	0·88	1·07
Vacuum cleaning ..	†	0·98	0·58	1·99**	1·17	1·29
Doors shutting ..	1·79*	1·49	1·58	1·65*	1·22	1·21
Footsteps	1·39	1·48	2·19**	2·48**	1·07	1·97**
Banging or hammering	3·57**	3·87**	3·94**	5·60**	4·69**	†
Fire being poked ..	1·93**	1·91**	1·66*	1·48	1·40	1·23
Piano	0·51	1·76*	1·26	0·70	†	1·00

Question 11(*a*): '*Do you hear.....from the flat (above, below, next door)?*'
Take the first noise and ask about above, below, next door, in turn. Then take the second noise and so on. *Only then* ask (*b*) for each noise heard, working across the table for each noise in turn.
(*b*) '*Does.....from the flat.....disturb you?*'

* Ratio is significant at 5 per cent level.
** Ratio is significant at 1 per cent level.
† Group-interviewer interaction significant.

Table III shows the results obtained for a battery of questions dealing with ten noises. The actual form and order of questioning is given below the table. There is plenty of evidence of interviewer variability. The largest values of the ratio occur for the noise which is probably least capable of definition, 'banging and hammering'. Partly perhaps for this very reason it was rated the most disturbing noise in another

question. The significant group-interviewer interactions are probab due to difficulty on the part of the interviewer in dealing with th complicated set of questions.

Some evidence of how easy it was to affect the answers to the questions was provided in another way. Different questionnaires we used on alternate interviews, one asking about the flat above first, th other about the flat below first. Comparison of the results for the tw questionnaires suggest that when a housewife was uncertain as to th source of a noise she attributed it to the first flat mentioned.

It is scarcely surprising to find that an interviewer who obtained high proportion of complaints from her Group I sample tended also t get a high proportion of complaints from her Group II sample. Fo the average number of disturbing noises obtained in answer to que tion 11 (Table III) the correlation coefficient is 0·48. We are led there fore to expect that the results obtained by an interviewer will reflec to some extent her own opinions.

A short questionnaire was given to the interviewers working in th Group I and II flats after they had completed their interviewing. On of the questions was: 'If you lived in a flat would you rather hear som noise or no noise at all from the other flats?' Of the twenty inter viewers nine wanted some noise and eleven no noise at all. An analys of variance showed a significant difference in the answers obtained b these two groups of interviewers to the two questions of Table II whic were examined, question 4 (too much noise comes through the ceiling and question 24 (noise from inside the building is very disturbing). is true that the above question was put to the interviewers after th interviewing and the answers might possibly be said to have bee affected by their experiences. This, however, seems an unlikely ex planation, and there seems little doubt that the results obtained in th survey were in some cases influenced by the interviewers' own opinion

As far as comparisons between the Group I and II flats were co cerned interviewer variability was of rather academic interest, sin the same twenty interviewers had equal interpenetrating samples f each group, and comparisons could be made independently of inte viewer. When, however, we came to examine the results for th Group III flats this variability caused more concern. For whereas w found that the questions which asked about the noises heard varied fo the three groups of flats in a way that might be expected from th measured sound insulations of the floors, when we came to consid disturbance and complaints the Group III tenants with the lowe insulation were found to complain less than the Group I tenants wit the highest insulation. We felt a little uneasy in making the assumptio that the original twenty interviewers used in London for Groups I an II were drawn from the same 'population of interviewers' as the twelv used in Scotland for Group III, particularly as in the first case th interviewers were all English while in the second they were predom nantly Scottish. Fortunately, as has been said, two interviewers worke

throughout the inquiry, and we were able to compare their results with the overall results. Their results showed similar trends to those for all the interviewers as can be seen from the example below:

	Average Number of Noises heard from Flat Above			Average Number of Disturbing Noises from Flat Above		
	Group I	Group II	Group III	Group I	Group II	Group III
All interviewers	5·8	6·2	6·8	1·7	2·3	1·1
Two common interviewers ..	6·2	6·7	6·8	2·1	2·5	1·4

Thus interviewer variability could not account for the unexpected results for Group III in questions dealing with disturbance.

Summarising the results for the noise survey, we may conclude that the factual items when adequately defined showed no significant interviewer variability; but the opinion questions dealing with noise showed appreciable variability which was related to the interviewers' own opinions.

4 *Sickness Survey*

The second example to be considered is taken from the Survey of Sickness. This survey commenced in 1944 and continued with a few gaps until the beginning of 1952. Sampling and interviewing were carried out by the Social Survey, while the General Register Office dealt with the classification of the illnesses and the analysis of the results. During the first fortnight of each month a different sample of adults was questioned about their illnesses in previous months. At the time with which we are concerned the sample names and addresses were drawn from the National Register. The sample was of a two-stage design, the first-stage units being administrative districts, i.e. County Boroughs, Boroughs, Urban Districts, and Rural Districts. After suitable stratification the sample of first-stage units was selected with probability proportionate to the populations of the districts. With such a method of selecting the first-stage units it is possible to select the same number of second-stage units, i.e. persons, from each of the chosen first-stage units or districts, while still preserving the same overall probability of selection for each member of the population.

In many cases only one interviewer would normally have been assigned to one district. Where two were assigned to a town the sample would normally be so divided that one interviewer dealt with one half of the town while the second interviewer dealt with the other half of the town. Thus differences between interviewers could not be separated from differences between districts. Nevertheless tabulations which had been produced for each interviewer on a number of samples suggested that the number of illnesses obtained was to some extent dependent on the interviewer, since an interviewer getting a relatively large number

of illnesses in one town on one sample appeared also to get a relativel large number when interviewing in another town on another sample

We decided therefore to attempt to measure the interviewer var bility. No money was available for a special experiment; consequent any measurements had to be incorporated into a normal month's inte viewing without interfering with the operation of the survey. Th number of interviewers available was also limited by the requiremen of other surveys taking place at the same time. However, we were ab to provide two interviewers for each of twenty-three urban administr tive districts, mainly in the Greater London area. For each of th twenty-three areas two interpenetrating samples were drawn. Ther were thus forty-six interviewer-assignments and these produced a tota of 888 interviews, or an average of just over nineteen per assignmen The interviewing took place in June 1950.

For a number of the questions put during the interview an estimat of the ratio, F, of the between-interviewer variance to the within interviewer variance has been calculated and is shown in Tables IV–VI F will be distributed approximately as the variance ratio with (23, 842 degrees of freedom.

Table IV gives first a number of classificatory items not dealing wit illness. In four of the five items the interviewer variability proved t be not significant at the 5 per cent level. The fifth case provides a

TABLE IV

Sickness survey—classificatory items—
ratios of between-interviewer variance to within-interviewer variance

Attribute or Variate Analysed	F
Number of persons in household	1·08
Age of subject	0·86
Whether married	1·07
Number of rooms in dwelling	1·46
Subject living 'at home'	1·86*

* Significant at 5 per cent level.

object lesson in what should not be done. This classificatory ite appeared on the questionnaire as follows without any question:

Subject where living	At home	1
	In institution, hotel	2
	As a boarder	3
	In rooms	4
	As a resident servant	5

Classificatory items can be treated like this when the precoded answe clearly indicate what is required, e.g. married, single, widowed divorced. When, however, the distinction between 'at home' and 'i

ooms' can only with difficulty be decided by studying the interewers' instructions, it is scarcely surprising to find an interviewer ifference. This particular classification had appeared on the questionaire ever since the beginning of the survey, though, as far as can be aced, it was never used in any analyses.

When we come to the answers to the questions concerned with sickess there is evidence of appreciable interviewer variability. The interew commenced with the introductory question:

1. 'How was your general health during April and May?'

then continued with:

2. 'Did you have any illness, ailment, poisoning or injury of any kind trouble with long-standing complaints during April and May?'

ny illnesses mentioned were then recorded.

The interviewer then read out the following check list, 'pausing after ch item or group of items':

'Have you had anything wrong in the way of colds, catarrh or nose nd throat trouble, or anything wrong with your eyes, ears, teeth, head ains, chest, heart, stomach, or indigestion, liver, kidneys, bowels, or nstipation, legs, feet, hands, arms, or rheumatism, skin complaints, fectious diseases, or anything wrong with your nerves?' (For women ly) 'Have you had anything wrong in the way of women's comaints?'

The interviewer then went on to establish for each illness whether it as present in each of the two months and the number of medical nsultations and amount of incapacity it caused in each month.

Some results for these questions are given in Table V.

TABLE V

Sickness survey—illness questions—
ratios of between-interviewer variance to within-interviewer variance

Attribute or Variate Analysed	*F*
Whether ill in either month	1·00
Total number of illnesses reported	1·99**
Number of illnesses before check list used	3·14**
Number of illnesses after check list used	2·33**
Number of medical consultations in April	0·92
Number of medical consultations in May	1·27
Number of days away from work or indoors—April ..	0·73
Number of days in bed—April	0·99

** Significant at the 1 per cent level.

Although the number of people reporting any illness in the two nths does not differ significantly between interviewers, the number illnesses does and to a marked degree. It is possible to distinguish

between illnesses mentioned in answer to question 2 before and af the check list was used. Both those mentioned before and those m tioned after the use of the check list show a significant variation.

There was some reason to expect that medical consultations a incapacity might show interviewer variability, since it has been sho elsewhere[2] that on the Survey of Sickness the number of medi consultations reported varied to some extent with the period of rec demanded of the informant. It seemed possible that the intervie might influence the informant's recollection of events. However, results in Table V show no such effect.

At this time the survey included one other question which requir a very considerable effort of memory on the part of the informant.

10. 'Have you ever had mumps?'

Subject's decision:	
Definitely Yes	5
Probably Yes	6
Probably No	7
Definitely No	8
Don't know	9

The answers to this question were used by Logan[3] to show that the was no evidence of association between diabetes among adults and history of mumps. He demonstrated also, as might be expected, th there was a considerable memory failure with advancing age. How ever, we did not find any significant variation between interviewers the number of 'Definitely Yes' answers ($F = 1{\cdot}44$). Nor did the ad tion of the 'Probably Yes' answers produce a significant variati Thus we have no evidence so far to show that the interviewer affe questions involving memory.

Table VI gives the values of the variance ratio for the illnesses wh divided into eighteen groups based upon the International Statisti Classification of Diseases, Injuries and Causes of Death,[4] which w used by the General Register Office in classifying the illnesses. I seven of the groups no analysis has been made as the number of ca was very small. Out of the remaining eleven groups two show interviewer difference at the 1 per cent level. The first of these 'Other digestive', which consists mainly of constipation. The seco is the large group, 'Ill-defined symptoms'. One further group, 'Oth defined illnesses', shows a difference at the 5 per cent level. With su a result it is not surprising to find that when 'ill-defined symptoms' a excluded the remaining total number of illnesses does not differ signi cantly between interviewers, F being 1·44.

Thus the difference between the interviewers lies mainly in t number of symptoms recorded. This result is rather similar to th reported by Cochrane *et al.*[5] for the medical histories of coalminers obtained by four medical observers. Differences were found mainly f

TABLE VI

Sickness survey—illness groups—ratios of between-interviewer variance to within-interviewer variance. The variable used is the number of illnesses

Illness Group	International Classification	*F*
1. Tuberculosis of lungs	001–008	†
2. Psychoneuroses, mental disorders ..	300–326	†
3. Eye affections	370–389	0·94
4. Ear and mastoid	390–398	0·98
5. Rheumatism	400–402; 720–727	1·13
6. Heart and arteries	410–456	†
7. Affections of veins	460–466	†
8. Colds, influenza	470; 480–483	1·16
9. Sore throat	472; 473; 510	†
10. Other respiratory	471; 474; 475; 490–502; 511–527	1·32
11. Dental disorders	530–535	1·54
12. Ulcer of stomach and duodenum ..	540; 541	†
13. Other stomach	542–545	1·12
14. Other digestive..	550–553; 570–587	2·10**
15. Diseases of skin..	690–716	0·66
16. Other defined illnesses	010–299; 330–369; 467; 468, 536–539; 560; 561; 590–689; 730–776	1·63*
17. Ill-defined symptoms	780–795	2·09**
18. Injuries	N800–N999	†

* Significant at 5 per cent level.
** Significant at 1 per cent level.
† Insufficient cases.

mptoms rather than diseases. It seems probable in our case that the mber of symptoms mentioned may well be determined by the neral attitude or 'bedside manner' of the interviewer. In addition e form of interview permits the informant to plunge into a lengthy scription of the illness, and it may well be that the definite illnesses e seized upon and recorded while some of the symptoms are not.

In considering the results obtained from the illness questions it is cessary at the same time to consider how the answers were classified. ne classification used was clearly designed for classifying the diagnoses ade by doctors rather than the symptoms described by laymen, hough provision was made for a group of sixteen different symptoms ll-defined symptoms' in Table VI). What our interviews produced

was a mixture of doctors' diagnoses as reported by the subject, s diagnoses, and symptoms. Similar illnesses may well be reporte different ways. Where a doctor has been consulted the illness ma reported as just the illness or as the illness plus the associated sympto where no doctor has been consulted it may be reported as just symptoms. One is then faced with the difficulty of deciding w constitutes a separate illness. The method adopted in this survey to count each item mentioned and little attempt was made to comb or eliminate symptoms. This will inevitably lead to a certain amo of double counting where both the illness and the associated sympto have been given. An alternative method would be to attempt to co bine groups of symptoms into illnesses and to eliminate symptoms illnesses already given. This would almost certainly do something decrease the differences between interviewers. However, such a meth of classification would require testing to ensure that it was reproduci and that one was not substituting a 'coder variability' for an 'int viewer variability'. These difficulties lead one to speculate whether form of questioning used could be improved. Would it be better deal entirely with symptoms in the form of a series of Yes/No questio

Since the number of illnesses showed a significant variation betwe interviewers it is not possible to say whether there exists a signific variation between districts; the arrangement of the interviewing d not permit this. However, for those groups of illnesses where the in viewer variability is not significant it is then possible to examine between-district variation. Of the eight such groups two showe significant between-district variation. They were 'Eye affections' 'Rheumatism'. It has to be remembered that all the administra districts were urban and that they were mainly in the Greater Lon area. For the total number of illnesses, excluding ill-defined sympto there is also a significant between-district variation.

Conclusions

Broadly speaking, questions can be divided into two classes, fac and opinion. With the first type, interviewer variability, when i found at all, usually takes the form of one or more interviewers hav misunderstood their instructions. With the second type it is doub whether the person setting the question always knows precisely what means by it. It follows that both interviewer and informant will uncertain as to its meaning and different interpretations will be upon the questions and answers. Hence it is not surprising that results are sometimes affected by the interviewers' own opinions.

With factual or clearly definable items, measurement of intervie variability is a poor substitute for measurement of the error on e individual observation. More may be learnt of the causes of errors recalls and postal checks, though these checks can seldom be compl Such checks also provide evidence of the existence of constant bi and errors common to all interviewers, whereas the measurement

terviewer variability does not do this. However, checks are much ss easily arranged and, apart from postal checks, are more expensive carry out.

The measurement of interviewer variability is easily arranged in rveys that are confined to a few localities, and there is little excuse r not arranging to measure it in inquiries such as the noise survey. such cases the number of interviewers does not have to be increased. ut when the sample would normally consist of a number of widely attered first-stage units with only one interviewer working in each strict, the measurement can only be made by either doubling the umber of interviewers, with consequent increase in briefing and avelling costs, or reducing the number of first-stage units, with consequent increase in sampling error. Thus with widely spread national rveys the issue is not so clear, particularly when the items to be udied are of a factual nature. If the latter is true there will, for most ems, prove to be no significant interviewer variability, and it is oubtful whether the additional expense is justified. Perhaps as a mpromise the interviewer variability might be measured for only a rt of the sample, say that part drawn from the conurbations. On the her hand, with national opinion surveys an appreciable variation tween interviewers is to be expected and an attempt should perhaps made to measure it, even if it means reducing the number of first-age sampling units selected. Unless one can allow for this variability e cannot be sure how far the relationship between the answers to o questions is affected by the interviewers.

In what has been said about interviewer variability, probability mpling has of course been assumed. If purposive or quota sampling used, a considerable amount of interviewer variability will be found en on factual items. It is introduced by the interviewers in the ocess of selecting the people to be interviewed, as was demonstrated ite clearly in the results obtained by Moser and Stuart.[6] Thus the mments made above should be taken to apply only to probability mpling.

REFERENCES

RAY, P. G. and CARTWRIGHT, A. (1953). *Noise in Three Groups of Flats with Different Floor Insulations.* Social Survey Report (unpublished).

RAY, P. G. (1955). 'The memory factor in social surveys', *J. Amer. Statist. Ass.*, **50**, 344.

OGAN, W. P. D. (1951). 'Mumps and diabetes', *Mon. Bull. Minist. Hlth Lab. Serv.*, 136.

he International Statistical Classification of Diseases, Injuries and Causes of Death, 1948. HMSO, London.

OCHRANE, A. L., CHAPMAN, P. J. and OLDHAM, P. D. (1951). 'Observers' errors in taking medical histories', *Lancet*, 5th May, 1007.

OSER, C. A. and STUART, A. (1953). 'An experimental study of quota sampling', *J. R. Statist. Soc.*, A, **116**, 349.

SOME SAMPLING PROBLEMS IN CONNECTION WITH ACCOUNTING RECORDS

STANLEY F. JAMES
Department of Economics, Nottingham University

In this article Mr James describes in considerable detail investigation into the application of statistical methods to examination of accounting records. He draws attention to special problems encountered when sampling highly sk population distributions such as are frequently found in type of work.

Introduction

There has recently been a considerable awakening of interest in t possible application of statistical techniques, in particular sampli methods, to accounting records. Although such records are essentia statistics of *variables*, usually measured in money terms, most of uses of sampling which have been discussed have considered them statistics of *attributes*, where regard is paid only to whether a particu entry is correct or erroneous (see, for example, Vance,[9] in particu pp. 12–13, where this approach is justified from the point of view of auditor). A few studies have indeed appeared where accoun records have been treated in the more obvious way as variables Neter,[5] Rowles,[6] Shartle,[7] and Trueblood and Cyert[8]) but emph here has been mainly on which type of sample is most efficient whe considerable amount of information on the population is alre possessed. Further problems are also likely to arise in connection v the estimation of population values from samples and of the accur of these estimates, and insufficient attention seems so far to have b given in this connection to the difficulties known to arise when the b distributions show deviations from normality as marked as must o be expected with accounting records. This study attempts to illustr these difficulties by reference to two distributions encountered practice.

Access to the figures presented here became possible through kind co-operation of a team of workers under Dr F. A. Wells from Department of Industrial Economics at Nottingham University, wh had persuaded certain firms in the area to make their records availa to a research enquiry into productivity, carried out on Counterp Funds. With one particular firm it became necessary to decide whet the actual distribution by physical size of completed orders (the f manufacturing only to order) had changed significantly from the pre-

the post-war period. This made necessary the collection of two mplete distributions of accounting records to answer this particular uestion, and at the same time it became possible to compare various sults and calculations from complete inspection of the figures with ose obtained by a sampling procedure.

llection of the Figures

Although the initial main interest was on physical size of order, it eedily became apparent that the various orders were of so excessively terogeneous a product that it was only possible to treat them at all reducing them to a common money basis, and then to deflate the st-war figures by some necessarily rough type of price-level index fore any actual pre- and post-war comparisons were made. In order define the time period exactly, it was decided to consider as falling thin a given period only those orders completed in their entirety and voiced during that time period—thus there would be considered as curring in the October 1952 to September 1953 period an order tered, say, in June 1951 and dealt with mainly in 1952, as long as e last delivery and invoice relating to that order occurred after 1st ctober 1952. Some arbitrary convention like this is obviously neces- ry in order to keep the collection of figures on an objective basis, even ough there may be (and were) paradoxical examples such as the ove. Finally, it was decided to limit the collection of figures to two ars only, and after consultation with the accounting staff of the firm question it was decided that October 1936 to September 1937, and ctober 1952 to September 1953 were sufficiently representative of the mediately pre- and post-war periods for more or less valid conclu- ns about the whole of these periods to be drawn from them.

Although all the records of the firm were available, it was neverthe- s found that considerable labour was involved in the actual extraction the figures required—an obvious eventuality with accounting records ich makes the possible use of sampling methods of interest here. us in the order book for a given financial year could be found all ders received in that year stated in physical terms (but without any lication as to total value of order), and also (usually) date of comple- n. In the invoice book could be found the money value and descrip- n of goods actually despatched to the ordering firm, but in general h no indication as to whether this formed the whole of the order, or y a part of a much larger order, as happened quite frequently. Thus money value of any completed order could only be found by eful reference backwards and forwards between these books (and, of rse, order books for one or even two years previously may have to examined when large orders taking many months to complete are olved).

The distributions obtained for the two years are presented below in ble I (the figures in the two columns headed 'Sample' are included e merely for simplicity of reference and will be dealt with later).

TABLE I

Distributions of orders by money value

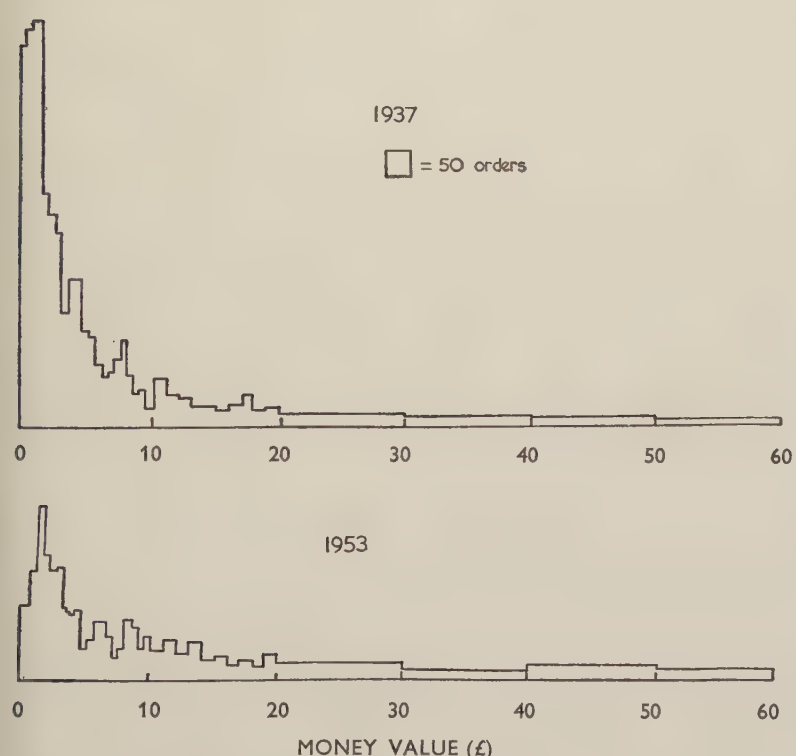
Range £	Midpoint £	1937 Distribution	1937 Sample	1953 Distribution	1953 Sample
0–0·5*	0·25	152	19	30	5
0·5–1	0·75	158	21	43	5
1–1·5	1·25	160	11	69	2
1·5–2	1·75	92	13	50	7
2–2·5	2·25	85	8	44	8
2·5–3	2·75	77	4	45	6
3–3·5	3·25	45	2	29	2
3·5–4	3·75	58	3	27	2
4–4·5	4·25	58	5	28	0
4·5–5	4·75	39	6	13	1
5–5·5	5·25	36	0	16	3
5·5–6	5·75	25	2	24	2
6–6·5	6·25	19	0	24	2
6·5–7	6·75	22	3	18	0
7–7·5	7·25	27	2	9	0
7·5–8	7·75	35	4	13	0
8–8·5	8·25	21	1	25	1
8·5–9	8·75	13	1	22	1
9–9·5	9·25	15	0	13	2
9·5–10	9·75	7	1	16	4
10–11	10·5	39	6	24	1
11–12	11·5	26	5	31	3
12–13	12·5	23	4	22	1
13–14	13·5	17	2	31	7
14–15	14·5	17	2	16	3
15–16	15·5	13	1	18	1
16–17	16·5	18	2	12	0
17–18	17·5	26	4	15	0
18–19	18·5	14	1	10	1
19–20	19·5	16	1	21	1
20–30	25	111	7	127	6
30–40	35	80	7	56	5
40–50	45	62	7	82	12
50–60	55	55	4	70	7
60–70	65	37	6	59	7
70–80	75	27	2	58	6
80–90	85	22	2	44	2
90–100	95	23	2	40	3
100–150	125	74	9	112	9
150–200	175	50	7	83	6
200–250	225	38	4	47	2
250–300	275	23	0	42	5
300–350	325	16	3	35	2
350–400	375	9	0	30	5
400–450	425	8	1	29	3
450–500	475	7	0	23	4
500–600	550	10	1	39	1
600–700	650	7	0	36	3
700–800	750	3	0	18	2
800–900	850	8	0	12	1
900–1 000	950	2	1	11	1

* To be exact, 0 to 9s. 11¾d., etc. The gain in simplicity of the method of descrip
adopted is worth the small inaccuracy involved.

TABLE I (*continued*)

Range £	Midpoint £	1937 Distribution	1937 Sample	1953 Distribution	1953 Sample
1 000–1 500	1 250	3	0	41(15)	3
1 500–2 000	1 750	25(1)	1	41(12)	8
2 000–2 500	2 250	3	0	8	0
2 500–3 000	2 750	0	0	6	0
3 000–3 500	3 250	1	0	1	0
3 500–4 000	3 750			1	0
4 000–4 500	4 250			1	0
4 500–5 000	4 750			0	0
5 000–6 000	5 500			0	0
6 000–7 000	6 500			1	0
Over 7 000	10 500			1	0
	Total ..	2 057	198	1 912	174

he frequencies for the first thirty-four groups are also shown graphic-ly (Fig. 1), with areas representing frequencies; only by means of a garithmic scale could the complete distribution be presented dia-ammatically, and this device would give a very distorted picture of e shape of the distribution.



. 1. Frequency diagrams of part of the distributions of orders by money value, showing the high degree of skewness. The long 'tails' are omitted.

G

Comparison of the Distributions

From a general glance at the two distributions obtained it car seen that both show in very marked degree the extreme skewness wl one must be prepared to expect in accounting distributions. T general shape of the distribution is fairly smooth, though each exhi a peculiar bump towards the upper end of the range, in the £1 5 £2 000 group for 1937 and in the £1 000–£1 500 and £1 500–£2 groups for 1953. (It should be noted that the group size varies c siderably, from 10s. at the lower end to £1 000 at the upper end device necessitated by the extreme skewness but which tends to obsc to a casual examination the smoothness which is in fact present.) T usual calculations give us the measures of central tendency and disp sion in Table II.

TABLE II

Characteristics of order distributions

	1937	1953
Total frequency	2 057	1 912
Mean	74·294	206·432
Variance	60 519	248 283
Standard deviation	246	498
Coefficient of variation ..	3·3	2·4

The reason for the appearance of the bump, and for its not mov along the curve as the level of prices increases, is one that concern immediately. Of the 25 orders in 1937 falling between £1 500 £2 000, 24 were from the same extremely large customer, regul repeating an order for a constant physical and money amount; si larly 26 orders in 1953 falling between £1 000 and £1 500, and falling between £1 500 and £2 000, were from this same custon regularly repeating an order, now for two varieties, for a const physical amount considerably lower than before but for more or the same total value. The appearance, then, of these particular bum and the paradoxical behaviour with respect to the price level, can b be explained by the ordering policy of one extremely important c tomer, a firm which instead of placing a single order within a year say, £40 000 of product preferred to give regular re-orders of a m smaller kind, and constant in value. For this reason these particu orders are omitted from consideration in the attempt to compare distributions of the two years, and the figures given in brackets in relevant groups are employed.

In any attempt to compare the distributions of the two years i obviously essential to eliminate the effect of changes in the gen

rice level of products in this period. The blind use of deflation by a rice index has certain pitfalls which cannot be avoided here; for xample, it obscures any possible change to a generally higher or lower uality of the product ordered in the periods under examination. nspection of the detailed order specifications during the collection of ie figures showed that there had been no very important shift of this ind. The price index used for deflation purposes was one specially alculated by the research project on productivity mentioned above; ith 1937 taken as the base year the index for 1953 is 288.

Using this figure (unfortunately not available until after the collec-on of the data) it remains to re-group the 1953 distribution into groups hose ranges correspond directly to those of 1937 inflated by 2·88. Vith only the grouped figures for 1953 available this means that fre-uently an even distribution of the ungrouped figures within the actual roups must be assumed; where this assumption becomes of importance, s in calculating 1953 frequencies corresponding to those in the £5–£10

TABLE III

Comparison of 1937 *and* 1953 *distribution by the* χ^2 *test*

Range (1937 values) £	1937 Actual	1937 Theoretical	Contribution to χ^2	1953 Actual	1953 Theoretical	Contribution to χ^2	Theoretical Relative Frequency
0–0·5	152	149·4	0·05	134	136·6	0·04	0·074
0·5–1	158	153·5	0·13	136	140·3	0·11	0·076
1–1·5	160	127·9	8·46	85	116·9	7·94	0·063
1·5–2	92	74·6	4·34	51	68·2	3·96	0·037
2–2·5	85	74·6	1·53	58	68·2	1·39	0·037
2·5–3	77	66·3	1·85	50	60·6	1·69	0·033
3–3·5	45	47·6	0·14	46	43·5	0·15	0·023
3·5–4	58	50·6	1·15	39	46·3	1·05	0·025
4–4·5	58	49·0	1·73	36	44·8	1·61	0·024
4·5–5	39	40·3	0·04	38	36·8	0·04	0·020
5–10	220	217·3	0·03	195	198·6	0·02	0·107
10–15	122	115·9	0·33	100	105·9	0·30	0·057
15–20	87	101·9	2·38	108	93·1	2·23	0·050
20–30	111	142·5	7·77	162	130·2	7·08	0·070
30–40	80	88·8	0·95	90	81·2	0·91	0·044
40–50	62	66·3	0·32	65	60·6	0·29	0·033
50–60	55	55·3	0·01	51	50·5	0·00	0·027
60–70	37	43·9	1·19	47	40·1	1·11	0·022
70–80	27	28·1	0·08	27	25·6	0·04	0·014
80–90	22	24·0	0·20	24	21·9	0·17	0·012
90–100	23	24·6	0·10	24	22·5	0·10	0·012
100–150	74	88·8	2·70	96	81·2	2·53	0·044
150–200	50	59·0	1·54	63	53·9	1·40	0·029
200–250	38	45·3	1·40	49	41·4	1·21	0·022
250–300	23	23·6	0·01	22	21·6	0·02	0·012
300–400	25	23·0	0·19	19	21·0	0·17	0·011
400–500	15	12·0	0·82	8	11·0	0·75	0·006
500–700	17	16·7	0·00	15	15·2	0·01	0·008
Over 700	21	23·0	0·00	19	19·1	0·00	0·011
Total ..	2 033	2 033·9*		1 857	1 856·8*		1·003*

* Error due to rounding off figures.

region of 1937, a change to wider groups has been made to elimi
any serious error that might arise from this assumption.

The 1953 frequencies having been obtained in this way, the hy
thesis was then made that the 1937 and 1953 distributions were the
selves drawings from the same theoretical distribution, the relat
frequencies of which were taken as the weighted average of the 19
and 1953 relative frequencies. This hypothesis was then tested by a
goodness of fit test, involving the calculations shown in Table III.

The resultant χ^2 gives a value of not quite 76, a figure highly sign
cant (1 per cent point = 56·9) with 28 degrees of freedom. Thus in sp
of the very close resemblance between the shapes of the distribution
the two years, the hypothesis that the basic distribution of size of orde
(measured in money terms) has remained unchanged is untenab
This is scarcely surprising if one considers the multifarious causes whi
must be present to lead to an order for a given amount being plac
with this particular firm during a given year and the probability
considerable changes in many relevant respects over the interveni
years. What is perhaps surprising is the extent of the agreement b
tween the two years, the only really large discrepancies occurring in t
£1–£1 10s. group, where there is a marked deficiency in 1953, and
the £20–£30 group, where there is a marked excess. In fact, althou
the χ^2-test shows significant differences between the two distributio
a closer inspection of the figures suggests that these differences are d
to changes in the *detail* of the distribution, such as must almost alw
be expected in such comparisons, and not *radical* changes in the sha
of the distribution of the kind which would be important for the p
ductivity enquiry. In other words, comparison of the complete distri
tions, where about 2 000 items are observed in each case, may be
precise for our purpose; it will be of interest to see firstly whet
samples can adequately reflect the shape of the complete distributio
and also whether comparison between two samples, with, say, about
items in each, will show as significant the detailed changes that oc
here.

The Characteristics of the Distributions

Before proceeding to discuss the problems of sampling from these t
distributions it is enlightening to consider in more detail the act
character of the distributions with which we have to deal, and
particular to compare their shape, when plotted, with the well-kno
symmetrical bell-like shape of the normal distribution. This is desira
because many of the simple statistical methods and tests assume th
the data considered come from this type of distribution, although
practice substantial differences between the actual distribution and t
ideal normal one may not seriously affect the tests.

For distributions with a single peak, divergences from the norm
shape may be of two main types—firstly, a given distribution may n
be symmetrical, but may have a lopsided appearance, when it is sa

exhibit *skewness*; secondly, even if symmetrical in appearance, the istribution may have a sharper or flatter peak than would appear in normal distribution; the degree of sharpness of the peak is called *urtosis*, a distribution with a sharper peak than the normal distribution eing called leptokurtic, while one with a flatter peak is called platy-urtic. Special measures have been devised for skewness and for kurtosis or details see the Mathematical Appendix, First Section), which take ne value of 0 if there is no difference in these respects from normality the distribution considered. The actual values for the two distribu-ons obtained here are:

1937: skewness 4·9; kurtosis 123·6

1953: skewness 3·7; kurtosis 182·9

ven to the casual eye these figures seem to be alarmingly different from ne ideal figures of zero in each case. The effect of large divergences om normality is considered at length in the Mathematical Appendix, econd Section, but may also be briefly indicated here. Thus the test may be employed to test whether two samples could have come om populations with the same variance; z is calculated as half the atural logarithm of the ratio of the two sample variances, and the ariance of z can be calculated from the size of the two samples. With ne degree of kurtosis found here, the variance of z is over 35–60 times large as if the distributions were normal.

he Problems of Sampling

It has been mentioned above that although all the records were vailable to provide the complete distributions, nevertheless consider-ble labour was involved in the extraction of the data. Hence a further roblem of considerable interest to us is to what extent samples instead complete distributions can be used to answer the questions with which e have been concerned, and at the same time reduce the labour in-olved to more manageable proportions.

The first problem that must be met is that, in Yates's terminology,[10] e lack a suitable frame. The order book itself, although it lists every der separately, gives no clue to the money size of the order; whereas e order book usually (but not always) gives the date of completion the order, this does not identify the last entry relevant to that order the invoice book more closely than to within twenty pages. Where date of completion is given in the order book, nothing further mains but a complete check through all the entries in the invoice ok before the required entries can be found and identified as pertain-g to that particular order.

It would thus seem much more simple to work from the invoice book the order book rather than the reverse. But if each entry in the voice book is given an equal probability of being drawn, this means at items representing partial completion of big orders are placed on

an equal footing with items representing complete small orders. If t former are rejected, it means that repeated sampling must be carri out until the requisite number of 'completion of order' items have be obtained, a process which may entail a wasteful amount of drawing a examination of items; if these 'partial completion' items are not re jected, and their money value is inflated to the money value of the to order on a proportional basis, this would mean that the bigger orde are being given a greater probability of being selected than the smal ones, but that it is not known what this probability is (there being easily deducible rule from the behaviour of the firm). The meth finally adopted was to carry out a ten per cent sample of the tw distributions by selecting at random ten per cent of the pages in t invoice book, and by taking into consideration only those entri referring to completion of orders or to complete orders in themselve This method can only deviate markedly from an ideal random samp if there is any connection between adjacent items on a page. Th would seem only to be the case when orders are completed at the sam time for the same firm; when this happens it is true that the orders a of more or less the same order of magnitude, but the actual frequen of such occurrences in this particular example seemed to be relative unimportant. The possibility of introducing non-randomness into t sampling in this way through choosing an operationally simple meth of selection must be borne carefully in mind, however, especially wh somewhat peculiar results are found in comparing samples and popu tions.

One further point as regards the collection of the samples must made. Whenever a 'completion of order' entry is found on a selec page, the money value given there is only that of the last batch of goo dispatched; reference back to the order book was made to find the s of the complete order, and the value of this was then calculated on proportional basis. This may in some cases lead to inaccuracy, es cially where the first delivery on this order included some initial c not to be found elsewhere. Experience in working with these figu suggested nevertheless that such inaccuracies would be small, and t proportional method is considerably easier and quicker to employ th the alternative of searching the invoice book for all deliveries made w respect to a given order, and totalling. The correct total in connecti with any partial completion entry was of course known exactly, owi to the work done on the collection of the complete set of figures; b in order to test the efficacy of the sampling techniques it is necessa all the while to proceed as if in ignorance of the exact figures, and use them solely to check up on the deductions made from the sampl

The figures obtained by this method of sampling are those given Table I, where for ease of reference they are given next to the figu for the complete distributions. Since the sampling technique employ here has not been an ideally random procedure, the first necessity is investigate whether the samples obtained give any indications of bi

or reasons which will appear later, it is best done in this particular case y means of a χ^2 goodness of fit test; for this, the population frequencies re taken as the correct ones and used to provide theoretical figures for ıe frequencies in the various groups of the two samples. The test self requires that coarser groups are employed than were in fact vailable, in order to avoid the appearance of very small frequencies. he calculations for 1937 and 1953 are shown in Tables IV and V spectively.

TABLE IV

Comparison of sample and population for 1937

Range £	Population Frequency	Relative Frequency	Actual Sample Frequency	Theoretical Sample Frequency	Contribution to χ^2
0–0·5	152	0·073 89	19	14·6	1·326
0·5–1	158	0·076 81	21	15·2	2·213
1–4	517	0·251 34	41	49·8	1·555
4–10	317	0·154 11	25	30·5	0·992
10–20	209	0·101 60	28	20·1	3·105
20–50	253	0·122 99	21	24·4	0·474
50–100	164	0·079 73	16	15·8	0·003
Over 100	287	0·139 52	27	27·6	0·013
Total ..	2 057	1	198	198	9·681

TABLE V

Comparison of sample and population for 1953

Range £	Population Frequency	Relative Frequency	Actual Sample Frequency	Theoretical Sample Frequency	Contribution to χ^2
1–2	192	0·100 37	19	17·5	0·128
2–10	366	0·191 32	34	33·3	0·015
10–20	200	0·104 54	18	18·2	0·002
20–50	265	0·138 52	23	24·1	0·050
50–200	466	0·243 72	40	42·4	0·136
200–500	206	0·107 74	21	18·7	0·283
Over 500	217	0·113 49	19	19·7	0·025
Total ..	1 912	1	174	174	0·639

The value of χ^2 obtained for 1937 is 9·681, which is well below the nificance level for 7 degrees of freedom, indicating that the sample completely satisfactory so far as this goes. For 1953, however, the lue of χ^2 is so small as to reflect an unusually good fit, for 99 per cent the time by chance the value of χ^2 is expected to be higher than 372! As far as this test goes, then, it seems that the two samples ovide representations of the actual distributions which are at least as curate as can be expected, so that the sampling techniques employed n be accepted on these grounds as introducing no important bias.

It should be noted further in connection with the samples that the mps which were obviously present in the basic figures are by no eans so here. Since it is necessary to use the samples as though

nothing was known about the exact distributions, all comparisons b
tween sample results and exact figures will be made for the n
adjusted distributions only.

Deductions from Samples

Samples can, as previously remarked, be used either to give appro
mate representations of complete frequency distributions, or to g
estimates, with an ascertainable degree of accuracy, of characterist
of these distributions. As a check on the reliability of the sampli
technique employed it has already been seen that the present samp
represent fairly faithfully the distributions from which they are draw
To follow up this point, it is interesting to see what conclusions wou
have been reached if the sample distributions had been used for t
comparison between the two years, as was done for the comple
distributions earlier. With the same approach, a similar table
Table III can be constructed, and the figures obtained are shown
Table VI.

TABLE VI

Comparison of 1937 *and* 1953 *samples*

Range (1937 values) £	1937 Actual	1937 Theoretical	Contribution to χ^2	1953 Actual	1953 Theoretical	Contribution to χ^2	Theoretical Relativ Frequen
0–1	40	38·32	0·073	32	33·68	0·083	0·19
1–2	24	18·10	1·925	10	15·90	2·191	0·09
2–5	28	28·21	0·002	25	24·79	0·001	0·14
5–10	14	12·24	0·252	9	10·76	0·287	0·06
10–25	31	34·06	0·275	33	29·94	0·314	0·17
25–50	18	19·16	0·070	18	16·84	0·080	0·10
50–100	16	15·44	0·021	13	13·56	0·023	0·08
Over 100	27	32·47	0·921	34	28·53	1·048	0·18
Total ..	198	198·00	3·539	174	174·00	4·027	1·01

Total $\chi^2 = 7{\cdot}566$

The value of χ^2 obtained from this test is 7·566, and with 7 degr
of freedom the 5 per cent value of χ^2 is 14·067; hence if the comparis
between the two distributions had been made solely on the strength
the two samples, the conclusion would have been reached that the
was no evidence of significant change. Referring back to the previo
remarks on this subject it can be inferred that with samples of this si
only fairly radical changes in the distributions would be brought
light, and since only changes of detail appear to be present in this cas
no significance is observed. Since for the productivity enquiry on
radical changes were of interest, the samples could provide adequate
the requisite information on distributions.

On the other hand, the samples might have been used, for exampl
simply to provide information on the means and variances of the tw
distributions, and it is in assessing the accuracy of these estimates th
particular care must be taken when the basic distributions sho

nsiderable deviations from normality. The following table enables a mparison to be made between the values obtained for samples and r distributions.

TABLE VII

Comparison of sample estimates with population values

	1937		1953	
	Distribution	Sample	Distribution	Sample
No. of observations	2 057	198	1 912	174
Mean	74·29	52·50	206·43	208·60
Variance	60 519	24 831	248 283	168 181

The immediately striking point about these figures is the extent to nich the samples tend to underestimate the population variances. nce in assessing the accuracy of the sample means knowledge of the riances is necessary, this point will be taken up first. It was men- ned earlier that very considerable adjustments should be made to the ndard statistical tests when great divergences from normality are esent in the data. The same warning applies when limits within nich the true value is expected to fall are attached to sample estimates. ven the sample variance as 24 831 for 1937 and 168 181 for 1953, e would expect the true value to fall, with 95 per cent probability, thin the following limits:

If the distributions are normal:

20 100–30 400 for 1937;

131 000–208 200 for 1953.

Allowing for the amount of kurtosis indicated by the sample values 71·9 and 10·0 for the two years:

7 100–86 600 for 1937;

104 900–266 700 for 1953.

us, in each case the true variance (60 500 for 1937 and 248 300 for 53) lies outside the range that would have been used if no attempt d been made to adjust for non-normality, but inside the range where ne adjustment had been made, even though this adjustment was ed on the sample values for kurtosis, and not on the true population ues. (For further details see the Mathematical Appendix, second tion.)

Similar difficulties arise in estimating the accuracy of the mean. For nples of our present size the mean can be assumed to be distributed rmally with variance $\sigma^2(1-f)/n$, where σ^2 is the population iance and f the fraction of the total population included in the nple. Again, however, it must be supposed that we only have

information from the sample, that $f = 0{\cdot}1$, $n = 198$, and that the o figure available for σ^2 is the sample value of 24 831. This would t give 95 per cent confidence limits of 31·3–73·7 for the true m although the population value is 74·3. It is obviously the san underestimation of the variance which causes the trouble, since w the correct variance the limits would be 19·0–86·0. The 1953 case d not present the same problems, as there the sample mean is so nea the true value that the latter falls well within the 95 per cent limit 143·5 to 273·7 even if the sample variance is employed; but it must remembered that the sample for this year showed a most unusua good fit to the complete distribution.

The simplest solution to the difficulty is to use for this purpose largest value for the variance as given by the 95 per cent limits cal lated from the adapted z-test (see above), though this is likely to me that wider limits are given than strictly necessary. For further deta concerning this matter the reader is referred to the Mathemati Appendix, second section.

Transformation of Data

An alternative approach to the problems concerning samples fr extremely non-normal distributions is to search for some transformati which will bring the distributions into more nearly normal for Again, since the purpose of sampling in cases of this kind is to av the labour of complete enumeration, it is of little avail advocat complicated methods of obtaining really adequate transformatic The simplest that seems possible for the present type of distributio a logarithmic one, and this was used for the complete and sam distributions for 1937. The new mid-point of the groups was take the mid-point of the transformed range rather than as the transfor value of the original mid-point in each case; for most of the gro there would be little difference between the two values. For the f group of all some special treatment is obviously necessary, and si the lowest recorded value was 2s. 3d., this group was taken as in t being from 2s. to 10s., transforming to -1 to $-0{\cdot}70$.

Even with this simple transformation a fair approximation normality is obtained. The measures for skewness are now reduced 0·125 and 0·003 for population and sample respectively, while t measures for kurtosis are $-0{\cdot}54$ and $-0{\cdot}86$. Of course, only mat matically derived tests can assess the importance of such values; h they confirm the immediate impression that significant skewness absent from the transformed distributions, but indicate that so appreciable kurtosis remains (see Appendix). Sufficient improveme is certainly achieved to enable comparisons between samples and co plete distributions, or between distributions for different years, to made in a simple way. Quite straightforwardly, the standard tests a comparisons can be made with the sample and population values mean and variance of the distributions of logarithms.

Also, it is possible to use the sample values of mean and variance—
322 and 0·955 respectively—as estimates of the values for the com-
ete distribution (in fact 0·896 and 0·946). Hence calculations, given
the Mathematical Appendix, third section, enable us to obtain
imates of the mean and variance of the original distribution, which
rn out to be: mean, 72·85; variance, 638 500. The true values, as
ven in Table II, are 74·3 and 60 519, so that this method gives a very
od estimate of the mean, but far too large a one of the variance. No
ry definite conclusions can be made on the strength of just one result
this kind. However, the indications seem to be that small departures
m normality in the distribution of logarithms, such as we have here
ith noticeable kurtosis but little skewness), will not have much effect
the final estimate for the mean, but may have a considerable effect
the variance. Thus the logarithmic transformation may prove of
e in some directions—as in facilitating comparisons between years,
.—but may cause additional complications in connection with the
oblem of obtaining reliable estimates from samples. (Again see the
rd section of the Mathematical Appendix for a further discussion of
s point.)

nclusion

It is not to be disputed that the use of sampling techniques can be
great value in connection with accounting records. This is especially
e where, for purposes of management or prediction, estimates may
required of particular aggregates which may prove to be exceedingly
ious to extract, owing to the particular book-keeping techniques of
firm. Often, too, these estimates may only be required very
proximately, so that the considerable labour involved in finding the
ct figure would be time wasted.

On the other hand, it is obvious that deductions from the samples
st be made with especial care, since the distributions to be met with
y often show far larger deviations from normality than are customarily
t with in most fields of statistical application. Where the samples
required merely to give some indication of the actual distributions
se difficulties do not arise, but they become particularly acute in
ng the samples to provide estimates for particular figures and in
essing the probable degree of accuracy of these estimates.

n case the difficulties which have been discussed seem so great as
make it preferable after all to attempt complete examination of the
uisite data, it may be of value to end with an indication of the
ing of time which sampling is likely to achieve. The figures for the
complete distributions given above were collected in about 120
rs. On the other hand, the figures for the two samples took only
ut eight hours to collect, owing partly to the rougher method of
culation used there, and the remaining work could all be performed
bout six hours by an experienced computer.

MATHEMATICAL APPENDIX

The Characteristics of the Population

The particular orders mentioned in the text were omitted for calc tions given below, and the frequencies given in brackets in Tab employed. With these adjusted frequencies the following figures the characteristics of the two distributions are obtained.

Table 1

	1937	1953	1953 adjusted for Price Chang
μ_1'	54·5	167·71	58·2
μ_2	27 670	201 820	24 330
μ_3	41 745 000	950 402 000	39 786 000
μ_4	96 933 × 10^6	7 572 965 × 10^6	107 205 × 1
β_1	82·261 9	109·880 9	—
β_2	126·603 7	185·924 9	—
Kurtosis	123·604	182·924 9	—
Skewness	4·889	3·730	—
Pearsonian criterion	3 156·0	64·36	—
Tests of normality:			
$\sqrt{\beta_1}$	9·061	10·482	—
g	0·428	0·452	—

In this table the usual notation is employed, so that

$$\mu_1' = \sum_{i=1}^{n} f_i x_i \Big/ \sum_{i=1}^{n} f_i$$

$$\mu_r = \sum_{i=1}^{n} f_i (x_i - \mu_1')^r \Big/ \sum_{i=1}^{n} f_i \qquad (r = 2, 3, 4)$$

$$\beta_1 = (\mu_3)^2/(\mu_2)^3$$

$$\beta_2 = \mu_4/(\mu_2)^2$$

The measure for kurtosis given is $\beta_2 - 3$, and that for skewnes Pearson's suggested formula:

$$\frac{\sqrt{\beta_1}(\beta_2 + 3)}{10\beta_2 - 12\beta_1 - 18}$$

The Pearsonian criterion also given is

$$k = \frac{\beta_1(\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)}$$

d indicates which type of Pearsonian curve is best suited to fit the tribution (see Elderton[1]). For the sake of interest such a curve was culated for the 1937 figures; here $k = 3\,156$, which is so large as to ggest a Type III curve, i.e. of form

$$f(x) = \frac{a^l x^{l-1} e^{-ax}}{\Gamma(l)}$$

h origin at the beginning of the curve (theoretically k should be inite). However, calculations give

$$l = 0{\cdot}049 \qquad a = 0{\cdot}001\,33$$

that the fitted distribution is asymptotic to the y-axis instead of owing a steep initial rise from zero; also with the form given, with gin at the commencement of the curve, the mean would be = 36·7, so to produce a mathematical form fitting the first moment 54·5, the Pearsonian curve would have to have its origin at approxitely $x = 18$. In view of these two difficulties it would seem that the arsonian system of curves is insufficiently flexible to represent the o present distributions with a mode very near the beginning and an remely long tapering tail.

As well as the Pearsonian measures of skewness and kurtosis it is uable to have in the table the two statistics commonly used to test mmetrical and symmetrical deviations from normality, respectively β_1 and Geary's ratio g of mean deviation to standard deviation. It bvious at a glance that the two distributions show great divergences n normality, but the inclusion of these statistics helps to indicate actual extent of these divergences. In fact values of $\sqrt{\beta_1}$ and g of size would indicate significant differences from normality even in nples of size 10–15 (see Geary[2], Geary and Pearson[4]), and the olished tables do not treat of any samples so small that these values uld *not* be significant.

The values for β_2 also help to warn us that sample estimates of the pulation variances may be far less reliable than with approximately mal distributions. Geary[3] gives the variance of z, where

$$z = \tfrac{1}{2} \log_e \frac{(S')^2}{(S'')^2}$$

$)^2$ and $(S'')^2$ are two estimates of the true variance σ^2 based on nples of size n' and n'' respectively) as:

$$\frac{\beta_2 - 1}{4} \left\{ \frac{1}{n'} + \frac{1}{n''} \right\}$$

s will be 35–60 times as great for the present distributions as for a mal population.

uctions from Samples

The adaptation of the z-test given above can be used to fit confidence ts to the true value of the variance, given the sample variance, but

information is required concerning β_2, and for the present purpos is obviously necessary to assume that we have available only s measures of skewness and kurtosis as can be derived from our sam The characteristics of the two sample distributions are found to b given in Table 2.

Table 2

	1937	1953
μ_3	29 655 000	191 530 000
μ_4	45 730 000 000	285 290 000 000
β_1	58·272	8·451
β_2	74·880	10·049
$\sqrt{\beta_1}$	7·67	2·91
g	0·45	0·61

From the last two rows it can be seen that both samples very d nitely exhibit asymmetrical and symmetrical departures from n mality; the 1 per cent values for the two statistics are approximatel

$$\sqrt{\beta_1}: \quad 0{\cdot}40 \quad (n = 198)$$
$$0{\cdot}43 \quad (n = 174)$$
$$g \text{ (lower limit)}: \quad 0{\cdot}763 \quad (n = 198)$$
$$0{\cdot}761 \quad (n = 174)$$

Thus the samples do give adequate warning that the assumptio normality is quite untenable, and that the standard method of set confidence limits must be corrected in some way; one possibility i use the sample values for β_2 in the formula given above for the varia of z, and this is the method used in the main text.

As regards confidence limits for the mean, the correct method sh strictly be to use the t-distribution corrected for non-normality. Ge provides formulae for doing this, but these are far too complicate be helpful for general work of this kind. To give some idea of general effect of non-normality, very rough approximations to Gea formula are (to the order of n^{-1}, omitting terms of the order of $n^{-\frac{3}{2}}$ higher):

$$\mathrm{E}(t) = -\tfrac{1}{2}\sqrt{\left(\frac{\beta_1}{n}\right)}$$

$$\mathrm{Var}\,(t) = 1 + \frac{8 + 7\beta_1}{4n}$$

where $t = \sqrt{(n)}(\bar{x} - m)/S'$

and $(S')^2 = S^2(1 - f)$, S^2 being the sample variance.

For samples from a normal distribution with samples of size 150–200 here, the customary distribution of t is virtually identical with a rmal distribution with zero mean and unit variance. Taking = 52·50, $m = 74{\cdot}29$, and $S'/\sqrt{n} = 10{\cdot}63$, we get a value of $t = -2{\cdot}05$, tside the 95 per cent confidence limits of $\pm$ 1·96, which, as we have n, could indicate that the sample mean for 1937 appears to be nificantly different from the population mean. However, if the pulation value of β_1, 82·26 for 1937, is used in the above formula:

$$\mathrm{E}(t) = -0{\cdot}326 \text{ instead of } 0$$
$$\mathrm{SD}(t) = 1{\cdot}738 \text{ instead of } 1$$

thout working out exact probabilities for this distribution (still nsiderably skew, so that the probabilities from the normal distribu- n are invalid), it can be seen that the value of t will no longer be tside the 95 per cent range.

e Logarithmic Transformation

The characteristics for the transformed distribution and sample are en in Table 3.

Table 3

	Population	Sample
μ_1'	0·896	0·822
μ_2	0·946 4	0·954 7
μ_3	0·133 98	0·001 544 6
μ_4	2·203 59	1·951 62
β_1	0·021 2	0·000 002 78
β_2	2·460	2·141
$\sqrt{\beta_1}$	0·145	0·001 7
g	0·824	0·838

t can be seen that even by the use of this crude transformation a approximation to normality is obtained. Any significant evidence symmetrical deviation has now disappeared for the sample (the er cent value of $\sqrt{\beta_1}$ is 0·40), but there is still some symmetrical iation, as the sample value is just above the upper 1 per cent limit for 0·833; there is thus a degree of humpedness in the distribution of $= \log_{10} x$, which is also exhibited by the sample measure for kurtosis − 0·86. However, it seems that the assumption that y' is normally ributed is not too far from the truth, and that its distribution can purposes of comparison be assumed to depend only on the two ameters of mean and variance, estimated from the sample as:

$$a' = 0{\cdot}822$$
$$b' = 0{\cdot}954\ 7$$

The question of estimating the mean of the original variable x assessing its accuracy is rather more complicated. If

$$y = \log_e x$$

is distributed normally, with frequency function

$$(2\pi b)^{-\frac{1}{2}} \exp\{-(y-a)^2/2b\}$$

then the rth moment about the origin of x is

$$\mathrm{E}(e^y)^r = \mathrm{E}(e^{ry})$$

$$= \int_{-\infty}^{\infty} (2\pi b)^{-\frac{1}{2}} \exp\left\{ry - \frac{(y-a)^2}{2b}\right\} dy = \exp\{ra + \tfrac{1}{2}r^2 b\}$$

$= \mathrm{M}(r)$, where $\mathrm{M}(t)$ is the moment generating function of the distri tion of y.

Hence $\mathrm{E}(x) = e^{a+\frac{1}{2}b}$

$$\mathrm{Var}(x) = e^{2a+2b} - e^{2a+b}$$

so that these two (and indeed any other) moments can be calcula simply from knowledge of a and b; it must be remembered that a and b refer to the distribution of natural logarithms of x, and so

$$a = a' \times \log_e 10 \quad = 1{\cdot}894$$

$$b = b' \times (\log_e 10)^2 = 5{\cdot}060$$

Using these values of a and b, and the assumption of exact norma for the distribution of y:

$$m_1' = \mathrm{E}(x) = 72{\cdot}85$$

$$m_2 = \mathrm{Var}(x) = 638\,500$$

as given in the main text.

The sensitivity of these estimates to sampling errors in the valu a and b should also be considered. One approach of course is to the sampling variance of m_1' as m_2/n, but this is of little use if estimate m_2 is very unreliable. Alternatively, if

$$m_1 = e^{a+\frac{1}{2}b}$$

then the error $\delta m_1'$ in this value due to small errors δa and δb is g by

$$\delta m_1' = e^{a+\frac{1}{2}b}(\delta a + \tfrac{1}{2}\delta b)$$

i.e. $$\delta m_1'/m_1' = \delta a + \tfrac{1}{2}\delta b$$

Now the sampling error δa can be assumed to be normally distribu with zero mean and variance b/n, whereas δb is very approxima normal with zero mean and variance $(c - b^2)/n$, where c is the fou moment about the mean of y. Thus, again approximately, $\delta m_1'/m_1'$ be assumed to be normally distributed with zero mean and varia

$$\frac{b}{n} + \frac{c - b^2}{4n}$$

ing a value of 0·062 48, or a standard deviation of 0·25. Thus
portionate errors of more than 50 per cent (± 36) in the value of
would be likely to occur with only 5 per cent probability; the degree
error likely to be involved is very much what the use of the *true*
iance of x led us to expect previously (± 33·5), but of course it was
lack of knowledge of this true variance that caused all the diffi-
ties.

A similar approach to the above for the estimate

$$e^{2a+2b} - e^{2a+b}$$

μ_2 gives a standard deviation of $\delta m_2/m_2$ of more than 0·7; this
licates that this method is likely to be of little value for estimating
true variance of x; even if the logarithms were exactly normally
tributed the proportionate error might be as much as 140 per cent
hich indeed means that the approximations used to reach this result
of little value), and the presence of kurtosis is likely to cause further
ccuracies. The transformation thus provides little help in estimating
variance unless the roundabout procedure of equating two possible
asures of the sampling variance of the mean is tried, i.e.

$$\frac{m_2}{n} = (m_1')^2\left[\frac{b}{n} + \frac{c - b^2}{4n}\right]$$

is gives in this case the extremely good value of 65 566, but the only
tification for it seems to be that errors arising from the fact that
$= 3b^2$, whereas $c = 3b^2$ with exact normality of logarithms, may be
minated.

REFERENCES

LDERTON, W. P. (1938). *Frequency Curves and Correlation.* Cambridge University Press.

EARY, R. C. (1936). 'Moments of the ratio of the mean deviation to the standard deviation for normal samples', *Biometrika*, **28**, 293.

dem (1947). 'Testing for normality', *Biometrika*, **34**, 209.

EARY, R. C., and PEARSON, E. S. (1938). *Tests for Normality.* Biometrika Office, London.

IETER, J. (1952). 'Some applications of statistics for auditing', *J. Amer. Statist. Ass.*, **47**, 6.

OWLES, B. M. (1953). 'Sampling techniques and LIFO', *Accounting Review*, **29**, 244.

HARTLE, R. B. (1952). 'How scientific sampling controls accuracy in invoicing', *Journal of Accountancy*, **98**, 167.

RUEBLOOD, R. M., and CYERT, R. M. (1954). 'Statistical sampling applied to ageing of accounts receivable', *Journal of Accountancy*, **97**, 293.

ANCE, L. L. (1950). *Scientific Method for Auditing.* University of California Press.

ATES, F. (1949). *Sampling Methods for Censuses and Surveys.* Griffin, London.

A TASTING EXPERIMENT*

NORMAN THEODORE GRIDGEMAN
Division of Applied Biology, National Research Laboratories, Ottawa

In this article Mr Gridgeman describes the design
analysis of a subjective assessment of the difference in fla
between certain types of jam.

Introduction

Sucrose is sweeter than glucose by a factor that has a perso
equation and whose usual value has been variously estimated to
in the range of about 1·2–2. Glucose, in the commercial form kno
as corn syrup or confectioner's glucose, possesses certain desira
qualities as an ingredient of jam. Apart from the matter of cost, th
is then an open question as to whether it is a good thing to use so
glucose in jam manufacture; in other words, how would the custon
—and in particular a customer accustomed to all-sucrose jams—re
to it? This question was recently posed in Ottawa, with spe
reference to a sugar mix of three parts sucrose and one part gluc
which mixture makes a jam that is certainly not obviously differen
taste from a standard jam. It was decided to assay the difference
a laboratory panel consisting of a few dozen local volunteers. The des
and analysis of this trial are the subject of the present paper.

The original plan called for four independent manufacturers (h
after referred to by the first four Roman numerals), each to m
parallel batches of jam from the same raw materials. One,
'regular' batch, was to contain sucrose only, and the other, the 'gluc
batch, was to differ from the first by a weight/weight replacemen
twenty-five per cent of the sucrose by glucose. This was to be d
with two kinds of jam: strawberry and raspberry. Thus 4×2
jam samples were expected. In the event one manufacturer did
prepare any raspberry jams, so there were fourteen samples to deal wi

Information was sought on two aspects of the regular : glucose co
parison, namely sweetness and preference. Forty-eight unselec
subjects (thirty-nine men; nine women) from the divisional staff w
recruited as tasters. Each was presented with small samples (ab
five grams in a glass dish) of the jams in the form of seven coded pa
and was asked to indicate on a check sheet which member of each p
he thought was the sweeter and which he preferred. All the pairs w
presented together, but the coding key was varied in such a manner
to exclude any overall bias in the order of presentation of the sev
jam-pairs or in the left : right disposition of regular : glucose wit

* NRC No. 3850.

TABLE I

Distribution of votes for sweetness and for preference among the forty-eight tasters

Strawberry I

Prf \ Swt	R	G	N	Σ
R	21	10	0	31
G	7	6	0	13
N	2	0	2	4
Σ	30	16	2	48

Raspberry I

Prf \ Swt	R	G	N	Σ
R	13	18	0	31
G	9	6	0	15
N	0	1	1	2
Σ	22	25	1	48

Strawberry II

Prf \ Swt	R	G	N	Σ
R	18	7	0	25
G	14	7	0	21
N	1	0	1	2
Σ	33	14	1	48

Manufacturer II did not supply raspberry jams.

Strawberry III

Prf \ Swt	R	G	N	Σ
R	9	11	0	20
G	16	9	1	26
N	0	1	1	2
Σ	25	21	2	48

Raspberry III

Prf \ Swt	R	G	N	Σ
R	11	9	0	20
G	13	11	0	24
N	1	1	2	4
Σ	25	21	2	48

Strawberry IV

Prf \ Swt	R	G	N	Σ
R	15	9	0	24
G	10	12	0	22
N	1	0	1	2
Σ	26	21	1	48

Raspberry IV

Prf \ Swt	R	G	N	Σ
R	15	11	0	26
G	10	9	0	19
N	2	0	1	3
Σ	27	20	1	48

Grand Total

Prf \ Swt	R	G	N	Σ
R	102	75	0	177
G	79	60	1	140
N	7	3	9	19
Σ	188	138	10	336

wt = Sweetness, Prf = Preference, R = Regular, G = Glucose, N = Neutral. Roman merals indicate manufacturer.

For example, with strawberry I, 10 persons voted the glucose jam as sweeter but pre- ed the regular jam.

pairs. Within that restriction the order was randomised. The tast was done in air-conditioned single-seater booths under neutral lig Unsalted crackers and drinking (or rinsing) water were available.

Results

A summary is given in Table I of the way in which the forty-ei persons voted with regard to sweetness and to preference. A 'neutral' answers turned up, i.e. no vote for either alternative. Actua in briefing the subjects we urged them to make a decision no ma how little confidence they might feel in so doing, but some defaul nevertheless. Only one subject recorded neutrals for all seven pa Overall, the glucose jams received slightly more than 40 per cent the decisions on both sweetness and preference, whereas the regu jams received over 50 per cent. In both instances the vote differer is large enough (the probabilities are beyond the 5 per cent point) suggest that small but real distinctions exist between the two kinds jam for these subjects. These are 'two-tail' probabilities. Howev taking account of our knowledge that the regular jams cannot be sweet than the glucose jams, we may subject the sweetness differer to a 'single-tail' test of significance, which will make for a sharp distinction. In fact the probability of a fortuitous difference equal or greater than that observed (in favour of the regular jams be sweeter) is less than 1 per cent.

Sweetness vis-à-vis Preference

There is no evidence of any general association between sweetn and preference. Take, for instance, the overall proportion of prefere for glucose jams: it was 42 per cent among the votes cast for the regu

TABLE II

Sweetness-preference association among the tasters who recorded no 'neutral' judgement

Number of Associations in Pair Contrasts (maximum = 7)	Observed Frequency	Expected Binomial Frequency
0	4	0·20
1	1	1·54
2	6	5·24
3	8	9·89
4	4	11·20
5	7	7·61
6	6	2·87
7	3	0·46
Total ..	39	39

am as the sweeter and 43 per cent among those cast for the glucose am as the sweeter. Inspection of the results of the seven individual pairs reveals some variation in this respect, but all of it is within the normal range of random fluctuation. This finding applies to the group, of course, and not necessarily to the component individuals, ome of whom in fact do associate preference with greater or less weetness. In this connection it is of interest to segregate the subjects according to the number of occasions (maximum 7) they expressed a preference for the sample that they also judged to be the sweeter. For his purpose we shall exclude the nine subjects who gave one or more neutral answers. In Table II the observed frequency distribution is compared with the binomial expectations on the hypothesis that the association of preference with sweetness is a random variable with $p = 145/273 = 0{\cdot}531$.

A test for goodness of fit yields $X^2 = 21{\cdot}0$ for 4 degrees of freedom after pooling of the terminal pairs of groups), which is well beyond he customary significance levels. (Throughout this paper X^2 will be used to distinguish the calculated statistics from tabular χ^2.) We therefore infer that some subjects specifically based their preferences on heir judgements of sweetness. The overall mean of 145/39 = 3·718 is lose to 3·5, showing that the individual associations cancelled out for he panel as an entity, i.e. there were evidently about as many subjects who preferred the sweeter as preferred the less sweet.

Personal Differences

With regard to the sweetness judgements considered by themselves, here is no evidence of heterogeneity among subjects. The same can be

TABLE III

Distribution of sweetness and preference votes, excluding the tasters who recorded 'neutral' judgements

Number of Votes-in-7 for Glucose Jam	Sweetness Frequency		Preference Frequency	
	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency
0	1	0·87	1	0·76
1	6	4·58	4	4·06
2	7	10·32	9	9·27
3	14	12·92	12	11·75
4	11	9·71	8	8·94
5	4	4·38	5	4·08
6	1	1·10	1	1·03
7	0	0·12	0	0·11
Total ..	44	44	40	40

said of the preference judgements. It is, however, more than likely t this is a consequence of the comparatively small size of the trial small differences between the contrasted jams. A larger trial greater replication might be expected to reveal differences at pre masked. As it is, the distributions of votes-out-of-7 are almost exa what would be expected if all the tasting had been done by one sub instead of being spread over many and if the seven jam pairs ha common sweetness difference. This can be appreciated by mere insp tion of Table III, in which the observed frequency distribut (excluding subjects who gave neutral answers) are compared with binomial expectations for a perfectly homogeneous panel whose pro bility of casting a decision for a glucose jam is 132/308 = 0·429 sweetness and 121/280 = 0·432 on preference. Goodness-of-fit t support the intuitive conclusion that there is no evidence of hete geneity: when the 0, 1 and the 5, 6, 7 vote-groups are pooled, giv 3 degrees of freedom, we find X^2 to be 1·95 for sweetness and 0·23 preference, both obviously below significance.

Differences between Manufacturers and between Fruits

The average distribution of sweetness votes between 'regul 'glucose', and 'neutral' for the seven pairs (see Table I) is 188/7, 13 and 10/7. If all seven pairs were indistinguishable as regards th sweetness differences, and if all subjects were of equal sensory acu we should expect the individual distributions (e.g. 30, 16, and 2 strawberry I) to conform to the average; the degree of conformity be judged by χ^2 with (7–1) × (3–1) degrees of freedom. Making appropriate test we find $X^2 = 13{\cdot}4$, which falls on the 0·33 poin χ^2 with 12 degrees of freedom, and again is not significant. As m be expected from the lack of evidence of differences between all se pairs, neither is there any evidence of differences between the types of jam, strawberry and raspberry. The same can almost be about differences between manufacturers except that, with regar the preferences, there is a suggestion of such a difference. For exam manufacturer I collected 62 preference votes for his 'regular' whereas manufacturer III collected only 40. A χ^2-test for heterogen between the four manufacturers yields $X^2 = 12{\cdot}3$, which ne achieves the $P = 0{\cdot}05$ point of χ^2 with 6 degrees of freedom, nar 12·6.

Differences in General: Another Approach

If we ignore the returns of those subjects who recorded one or m 'neutral' votes, we can test for heterogeneity among the remainder means of Cochran's criterion[1] Q, whose limiting distribution is tha χ^2. This is a more searching criterion than those used in the preced sections, although here partially offset by our having to jettison so of the information. The criterion is in principle applicable to any b of data that can be set out in a complete table of M rows time

olumns, each entry being 0 or 1, expressive of some simple alternative. Let T be any row total, let U be any column total, and let $\Sigma T = \Sigma U = S$, the grand total. Then

$$Q = \frac{(N-1)\,[N(\Sigma U^2) - S^2]}{NS - \Sigma T^2}$$

ssessed against χ^2, with $N - 1$ degrees of freedom, is a criterion of the eterogeneity among columns of the within-column distributions of 1's. The same expression, with N replaced by M, and with U and T ransposed, yields another Q that, assessed against χ^2, with $M - 1$ legrees of freedom, is a similar criterion of row differences. Applying hese methods to the data at hand we obtain the results given in Table IV, in which it may be seen that only one Q, that of preference

TABLE IV

Probabilities of homogeneity: (i) among the sensory discrimination of the tasters; (ii) among the differences in the pairs of jams

		Sweetness	Preference
No. of tasters		44	40
No. of jam-pairs		7	7
Among tasters	Q ..	44·4	42·4
	P ..	0·43	0·34
Among jam-pairs	Q ..	7·5	14·3
	P ..	0·28	0·027

Q is Cochran's index, whose distribution is approximately that of χ^2.

ifferences among the seven jam pairs, is suspiciously high. This links vith the suggestion emerging in the previous section that there may be ome difference between manufacturers as far as these batches of jam re concerned. Whether this would be borne out by new batches is roblematic.

Discussion

In interpreting the results we must take into account the fact that ney stem from a laboratory sensory-difference trial, involving 336 aired comparisons. Even under conditions of such stringency only a mall difference emerged. It is accordingly unlikely that the public at arge, in normal (non-comparative) jam-eating practice, would sense ny difference between regular and glucose jams in so far as the samples sed in the present investigation are typical samples.

An interesting point emerging from the trial is the absence, overall, f any correlation between sweetness and preference, and the fact that his is mainly due to a cancelling out of individual correlations. For ome subjects did associate sweetness and preference, but there were

about as many who favoured the sweeter as favoured the less sw (according to their own taste decisions). This finding is, of cou conditioned by the particular magnitude of the sweetness differe Clearly, if one jam was of normal sweetness and the contrasted jam unsweetened, few would prefer the latter. Moreover, the cancelling referred to is applicable only to the population represented by t panel of adults, drawn from a group of youngish scientific workers, it would be presumptuous to generalise further.

Acknowledgements

The author would like to thank the Department of Agriculture, a the Food and Drug Directorate of the Department of National Hea and Welfare, for permission to use the experimental data.

REFERENCE

[1] Cochran, W. G. (1950). 'The comparison of percentages in matched sampl *Biometrika*, **37**, 256.

ARE CORRELATIONS ANY GUIDE TO PREDICTIVE VALUE?

ROBERT FERBER
University of Illinois, Urbana

When faced with the choice between various hypotheses purporting to explain the fluctuations of a certain variable, one is tempted to choose the hypothesis that shows the best correlation with known results in given circumstances and to use it to predict future fluctuations. In this article Dr Ferber examines the adequacy of this procedure in practice.

THE coefficient of correlation (r) is a widely accepted measure of the 'goodness of fit' of a line fitted to a set of observations. The higher is r, the better is the goodness of fit of the fitted line, once allowance has been made for degrees of freedom lost in the process of fitting the line. As a result of this correspondence r has come to be accepted, at least in economic and marketing studies, as a major criterion of the adequacy of a hypothesis. As between two alternative hypotheses attempting to explain the fluctuations of a particular variable, the one yielding the higher value of r tends to be accepted as the more adequate. Instances of this practice abound in the literature, and for this reason it does not seem fair to single out one or two cases as illustrative examples.

True, many statistical textbooks warn against this practice to the extent that the reader is cautioned against extrapolating fitted functions beyond the range of the observations or applying the functions to a situation where a change in basic structure is evident. Faced with the exigencies of practical work and in the absence of any simple alternative, however, even textbook authors tend to succumb to this simple expedient of selecting functions for predictive purposes on the basis of the values of their coefficients of correlation. From a theoretical point of view it can be shown that this practice is not necessarily valid. At the same time, however, it cannot be (or has not as yet been) proved that it is *not* valid. Under such conditions the crucial question becomes: how successful is this widely practised procedure from an empirical point of view?

This paper will focus, and present evidence, on the adequacy or otherwise of the coefficient of correlation as a criterion of predictive accuracy. (Throughout this paper r refers to the value of this statistic in the period of observation, the period whose observations form the base for fitting the particular function.) I begin by summarising the original study which led me to examine this problem. I then give an account of some further empirical tests designed to explore various aspects of the object, and conclude with a discussion of the implications of the results to empirical work.

The Original Study

In the original study[1] seven different forms of aggregate ann savings functions were fitted by least squares methods to one or m of three different periods of observation—1929–40; 1923–40; 1923 1935–40—using different units of measurement. In some cases economic variables were expressed in terms of aggregates at curr prices, in others as aggregates at constant prices, in others on a *capita* basis at current prices, in still others on a *per capita* basi constant prices. Annual data were employed in all cases.

The accuracy of prediction of these sixty functions was tested using each to predict the appropriate value of annual savings in 19 49. Computation of the average absolute percentage error (A) of th predictions for this three-year period and comparison with the c responding values of r^2 served as the basis for evaluating the adequ of the latter statistic as an indicator of predictive accuracy. (T coefficient of determination, r^2, was used instead of the coefficient correlation because all correlations were of the same sign, positive, a because of the greater meaningfulness of the former. Expressing, a does, the proportion of the total variance explained by the independ variables, the coefficient of determination yields a more realistic indi tion of goodness of fit than the misleadingly high values yielded by square root, especially in the eyes of the uninitiated.)

The relationship between A and r^2 (adjusted for varying degree freedom) for all sixty functions was found to be positive. Howe this tendency for rising predictive accuracy to be associated v declining goodness of fit was largely spurious, owing primarily to combination of data relating to different periods of observation. A group, functions fitted to 1929–40 yielded the highest values of r^2 exhibited relatively low predictive accuracy in 1947–49 (and in 19 49). Functions fitted to 1923–40 had lower values of r^2 but impro predictive accuracy; and functions fitted to the discontinuous per 1923–30, 1935–40—a period which omits the main depression yea had very low values of r^2 but were among the most accurate in postwar predictions. Hence the scatter diagram of all sixty pair observations[2] pictures, in effect, three separate constellations of po proceeding from the lower left to the upper right-hand corner of chart, thus accounting for the positive correlation.

At the same time, however, absence of relationship between r^2 an was shown to exist even when the period of observation was h constant, for each of the three periods of observation used. A sim lack of relationship was demonstrated when the average absolute p centage error in the period of observation was substituted for r^2. It also found that r^2 was of no value in differentiating between adequacy of two or more functions explaining the fluctuations of same dependent variable, when the amplitude (the relative rang variation) of this variable changes from one case to another, whet because of the use of different periods of observation or different u

f measurement. From a theoretical point of view these points are asily established; but in practice the plausibility that such relations nay nevertheless exist, and have not been shown *not* to exist, appears o have led to considerable misuse of the coefficient of determination in uch cases.

These findings are, however, subject to three major limitations, apart rom the obvious one of being based on only one set of data. These mitations are:

1. A change in basic structure may have occurred between the pre-var years and the early postwar years invalidating the applicability of ne prewar-derived functions for postwar prediction. If so, there is no eason why r^2, or any other statistic relating to the period of observation, nould be a reliable indicator of predictive accuracy. Such a change in asic structure may have occurred because such early postwar years as 947–49 were for consumers a period of adjustment from the restrictions laced upon them by wartime shortages and price controls, and there-ore may not have been representative of consumer behaviour under nore 'normal' conditions.

2. The postwar data on savings, income, and related variables are ubstantially outside the range of the data in the (prewar) periods of bservation. Although no change in the basic economic factors affecting ggregate savings functions may have occurred, the form of the savings unction may be altered in this higher range, e.g. the relationship may e linear for the prewar range but curvilinear at the higher postwar evels. From a statistical point of view this limitation is much the same s the first one—that because of changes in the parameters of the unctions there is no reason why r^2 should be a reliable indicator of redictive accuracy. They are therefore treated together, although nese limitations are not the same from the economic point of view.

3. The tests did not conform exactly with the hypothesis advanced to xplain the results. Even when the period of observation was held onstant, the amplitude of the dependent variable, savings, did not emain the same because of the different units of measurement used—ggregates in some cases, *per capita* units in others, adjusted for price nanges in some but not all instances. The poor performance of r^2 might onceivably be entirely due to these differences in amplitude.

The following pages present the results of a number of additional sts carried out on the same and on other data in an attempt to deter-ine the effect of removal of these three limitations on the foregoing ndings.

dditional Tests

The danger of the relationship between r^2 and A being influenced by change in basic structure—which will be our main concern in these sts—can be reduced by using prewar-fitted functions to predict the ependent variable either in those postwar years where the prewar attern appears to have been re-established or in prewar years for

which the same relationship is believed to have existed. Both
alternatives are employed.

In the case of the postwar predictions two series of tests were
ducted, one with an entirely different set of data. Two different
were also carried out based on prewar predictions, though both
based on the savings-income relationships.

Postwar Predictions: Savings Functions, 1951–52. Correlation coeffici
between r^2 and A for the same savings functions mentioned previo
(p. 114) were computed, using for A the average absolute percen
error for 1951–52. These years are sufficiently far removed from W
War II to reduce greatly any effects on consumer behaviour due to
war. In addition, the average errors in the predictions of savings
those years based on the prewar-fitted functions are in the same rang
generally between twenty-five and fifty per cent—as are the resid
in the period of observation, thus providing an additional basis
believing that for those years at least the prewar relationship had b
largely re-established. (On the other hand, it deserves to be mentio
that the same tendency towards overestimation of savings was appa
in 1951–52 as in 1947–49.)

To maintain the amplitude of the dependent variable constant th
correlations have been computed separately for each type of savi
variable as well as for each period of observation. The results, sh
in Table I, indicate the absence of any correlation between r^2 an
about as conclusively as before. Only three out of fourteen correlat

TABLE I

Correlation between r^2 and A for 1951–52
Predictions of aggregate savings functions

Period of Observation	Savings Variable*	Number of Functions	Square of Coefficient of Correlation between r^2 and A†	Significance Coefficient 0·05 Level
1929–40	S	5	(−) 0·59	No
	S/P	5	(−) 0·12	No
	$S/\mathcal{N}P$	4	0·01	No
	All‡	17	(−) 0·18	Yes
1923–40	S	5	(−) 0·29	No
	S/P	5	0·00	No
	$S/\mathcal{N}P$	4	(−) 0·70	No
	S/Y	6	0·55	No
	All‡	22	0·00	No
1923–30, 1935–40	S	5	(−) 0·79	Yes
	S/P	5	0·00	No
	$S/\mathcal{N}P$	4	0·43	No
	S/Y	6	0·82	Yes
	All‡	21	0·00	No

* Meaning of symbols is as follows: S = savings; P = prices; $\mathcal{N}$ = population; disposable income.

† A negative sign in parentheses preceding a coefficient indicates negative correla

‡ Includes functions whose savings variable was $S/\mathcal{N}$.

the table are significant at the 0·05 probability level, and the highest of ese is in the wrong direction, i.e. the correlation for functions explain- g the variations in the savings-income ratio for 1923–30, 1935–40, here a high goodness of fit was associated with a high error of predic- on! To be sure, the number of observations for each of these correla- ons is fairly small, but the general agreement of the results would em largely to counteract this limitation.

Postwar Predictions: Railroad Shippers' Forecast Functions, 1948–52. In e course of a recent analysis of the accuracy and structure of US ilroad shippers' carloadings forecasts[3] a number of regression func- ons were fitted to prewar (1927–41) data in an attempt to explain the recasts in terms of recent trends in shipments and in related variables; l the independent variables were lagged. These attempts were on the hole quite successful. The coefficients of determination were generally the order of 0·9 and above; the coefficients of most of the variables oved to be statistically significant at the 0·05 significance level; the solute values of the residuals deviated, on the average, within three r cent of the shippers' forecasts; and the use of the functions to predict e shippers' forecasts in the postwar period yielded relative errors of e same magnitude.

An additional virtue of the data from the present point of view is that e postwar range of observations is well within the prewar range. The aximum aggregate quarterly carloadings figure for 1948–52, 8·8 illion carloadings, falls below the more than 9 million carloadings ipped in several quarters during the late 1920's and is only slightly eater than the 8·2 million carloadings shipped in the last quarter of 41.

Three groups of regression functions for these data were available to mpute the correlation between r^2 and A: (*a*) eleven functions reflect- g a like number of alternative hypotheses propounded to explain the riations in the shippers' forecasts in the aggregate, (*b*) twenty-one nctions satisfactorily explaining the variations in the shippers' fore- sts for five industries, with four to five functions derived for each dustry, (*c*) twelve functions satisfactorily explaining variations in recasts of individual industries within selected regions. (The exact nctions used in these tests are listed in Tables 14, 20, and 29 of ference 3.) The values of A for these tests were based on the quarterly edictions of these functions for 1948–52, a period during which the curacy of the predictions was about the same as during the prewar riod.

The results of these correlations, shown in Table II, provide much re support for the existence of a relationship between r^2 and A than s been obtained so far. A highly significant negative correlation was tained for the total carloadings functions—the dependent variable re is identical for all of these functions. Two of the five sets of dustry functions exhibited close, and negative, correlations between and A. On the other hand, the correlation for another set of industry

TABLE II

Correlation between r^2 and A for 1948–52

Predictions of railroad shippers' forecasts

Level of Aggregation	Subdivision	Number of Functions	Square of Coefficient of Correlation between r^2 and A*	Significance of (Unadjuste Coefficient a 0·05 Level
National ..	—	11	(—) 0·85	Yes
Industry ..	Iron and steel	4	(—) 0·99	Yes
	Cement	5	0·00	No
	Flour	4	0·85	No
	Lumber	4	0·99	Yes
	Agricultural implements	4	(—) 0·94	Yes
	All	21	(—) 0·20	Yes
Industry within region	—	12	0·00	No

* A negative sign in parentheses preceding a coefficient indicates negative correlatio

functions (lumber) was also statistically significant, but positive. the industry functions as a group, however, a significant, though large, negative correlation was obtained. Here the amplitude of dependent variable differed from one industry to another. Howe in the other instance where the amplitude of the dependent vari varied from one function to another, the region-industry function correlation was detected between r^2 and A.

Prewar Forecasts: Aggregate Savings Functions, 1941. The perio observation for the aggregate savings functions previously descr ended in all cases with 1940. The possibility therefore presents its using these functions to predict savings, or the savings-income r in 1941 and correlating the resultant errors with r^2. Since part of was affected by American entry into World War II, savings in onl first half of the year were predicted.

As before, correlations were computed separately for each peri observation and for each type of savings variable in order to hold stant the amplitude of the dependent variable. For two of the t periods of observation no relationship was detected between r^2 an (Table III). For the discontinuous period of observation, however, correlation is in the right direction (negative) in all instances an statistically significant at the 0·05 probability level for three of the correlations. Except for two of the four cases where the amplitud the dependent variable was constant, however, the relationship is a very close one.

Prewar Forecasts: Aggregate Savings Functions, 1931–34. The fitting number of the aggregate savings functions to the discontinuous pe

TABLE III

Correlation between r^2 and A for first half of 1941
Predictions of aggregate savings functions

Period of Observation	Savings Variable*	Number of Functions	Square of Coefficient of Correlation between r^2 and A†	Significance of (Unadjusted) Coefficient at 0·05 Level
1929–40	S	5	0·00	No
	S/P	5	0·25	No
	$S/\mathcal{N}P$	4	0·00	No
	All‡	17	(—) 0·17	No
1923–40	S	5	0·00	No
	S/P	5	(—) 0·09	No
	$S/\mathcal{N}P$	4	0·89	No
	$S/\mathcal{Y}$	5	(—) 0·25	No
	All‡	22	0·00	No
923–30, 1935–40	S	5	(—) 0·90	Yes
	S/P	5	(—) 0·18	No
	$S/\mathcal{N}P$	4	(—) 0·08	No
	$S/\mathcal{Y}$	5	(—) 0·87	Yes
	All‡	21	(—) 0·57	Yes

* Meaning of symbols is as follows: S = savings; P = prices; $\mathcal{N}$ = population; $\mathcal{Y}$ = posable income.
† A negative sign in parentheses preceding a coefficient indicates negative correlation.
‡ Includes functions whose savings variable was $S/\mathcal{N}$.

observation, 1923–30, 1935–40, although originally done in order to dy the interaction of omission of depression years with postwar edictive accuracy of various types of functions, makes available for s analysis the data for the omitted years. Accordingly, for the enty-one functions fitted to this period, correlations were computed tween r^2 and A for 1931–34. The values of the coefficient of deter-nation (i.e. the square of the correlation coefficient) for each type dependent variable separately, and for all the functions, are as lows.

Dependent Variable	*Coefficient of Determination*
S	0·00
S/P	0·00
$S/\mathcal{N}P$	0·15
$S/\mathcal{Y}$	(—) 0·86
All	(—) 0·01

Only the coefficient of determination for $S/\mathcal{Y}$ is statistically significant the 0·05 probability level. In the other instances no relationship tween r^2 and A is apparent. It should be noted, however, that for s test the assumption of no change in basic structure may not be able. The relative errors of the predictions for 1931–34 were several es those for the period of observation and the values of the economic iables for these years were outside the range of the observation

(though the former phenomenon may be due solely to the near-values of savings in 1931–34).

Implications of the Results

The results of these additional tests of the relationship between r^2 A would seem to indicate that in comparing functions fitted to de dent variables of differing amplitudes such a relationship often doe exist. It might be noted that similar results were obtained in a few instances when the average absolute margin of error in the perio observation was employed as the measure of goodness of fit instea r^2. Much the same state of affairs is evident in an earlier study w a number of regression equations were derived in an attempt to exp fluctuations in employment and in consumption in various secto the economy for 1919–32 and were then used to predict the 193 figures.[4] Scatter diagrams of the relationship between the stand error as a percentage of the mean of the dependent variable in the pe of observation and the average relative margins of error in the ex polated years for between 63 and 137 of these functions, the num depending on the year for which extrapolations were made, revea almost complete absence of correlation (p. 15 of the reference).

From a rigorous point of view the foregoing results might be q tioned on the grounds that significance levels for the correlations computed on the assumption of a joint normal distribution of variables involved, together with random selection, an assump which would seem particularly doubtful in the present case. For reason the significance of the relationship was re-computed in a nu of instances, based on the coefficient of rank correlation. In instances—the 1929–40 functions in Table I; the functions for nati iron and steel, flour, all industries, and industry within region in T II; and the 1923–30, 1935–40 functions in Table III—the relation originally found to be non-significant at the 0·05 probability level also non-significant by this test. In addition three of the relation previously found to be statistically significant—iron and steel in T II, and the 1923–30, 1935–40 fitted functions for S and S/Y—we longer statistically significant at the 0·05 probability level.

Now, what do these results imply with regard to the pra problem of selecting functions (hypotheses) for predictive purp Clearly the coefficient of correlation is not a reliable measure for su purpose, nor does there appear to be at the present time any alterna single statistic adequate in this respect relating to the period of obse tion. Unsettling as it may seem, there does not appear to be statistical substitute for *a priori* consideration of the adequacy of basic hypothesis underlying a particular function. This is not to that statistical measures relating to the period of observation are o use in this regard. Determination of the statistical significanc coefficients, examination of the pattern of the residuals, and stud the contribution of each variable to the overall relationship ar

onsiderable value in establishing the validity of a particular hypothesis. uch analyses are particularly helpful in weeding out inadequate ypotheses. But when it comes to differentiating between the functions ıypotheses) that are left—the situation encountered in the studies ɔrming the bases for this paper—such statistics are no substitute for *priori* considerations. In fact, reliance upon goodness of fit in this egard may well bias the researcher towards those factors that will rovide a good fit in a particular situation and away from the more asic factors influencing the variable under study. This appears to be ue even where goodness of fit is used only to differentiate between ınctions the coefficients of correlation of which differ in a statistically gnificant sense. The procedure, advocated by some, of using the oefficient of correlation for this purpose is likely to produce fallacious esults. The fact that the coefficient of correlation of one function xceeded in a statistically significant sense that of another function—frequent occurrence in economics studies—provided little evidence ı this study in favour of the former function for prediction purposes. As for the coefficient of correlation itself, these results would indicate ıat its usefulness is limited essentially to its original meaning, a ıeasure of the goodness of fit in the period of observation. Even in the ıtter respect, however, the measure must be used with caution, for if ıe amplitude of the dependent variables is not the same, differences etween coefficients of correlation may provide no indication of the elative magnitude of the residuals of the various functions.

cknowledgements

I should like to thank Mr C. R. Blyth and Mr V Lewis Bassie for eading this paper and for helpful suggestions. I also wish to acknowdge the competent assistance of Miss Ramona Jean Rogers and Miss hyllis Barnard with the statistical analysis.

REFERENCES

Ferber, R. (1953). *A Study of Aggregate Consumption Functions.* Technical Paper No. 8. National Bureau of Economic Research.

Op. cit., Chart 4, p. 49.

Ferber, R. (1953). *The Railroad Shippers' Forecasts.* Studies in Business Expectations and Planning, Bureau of Economic and Business Research, University of Illinois.

National Resources Committee (1938). *Patterns of Resource Use.* US Government Printing Office.

A 2^4 FACTORIAL EXPERIMENT IN FOUR BLOCKS OF EIGHT: A STUDY IN RADIOACTIVE DECONTAMINATION

MARTIN K. BARNETT and FRANK C. MEAD, Jr.
Mound Laboratory, Monsanto Chemical Company, Miamisburg, Ohio.*

The exploration of complex practical situations is facilitate the prosecution of carefully designed factorial experime Such experiments provide conclusions of maximum genera and reliability at minimum expense. The authors illust this by a study of the effect of variations in four opera factors on the efficiency of a decontamination process removal of radioactivity from liquid wastes, and show statistically significant interactions were explained by rea able physico-chemical mechanisms.

Introduction

Among the complex experimental designs available for induc experimentation, the factorial design, i.e. one including all poss combinations of the independent variables at their chosen levels, se as arch-type. This is because the factorial experiment not only rev all primary effects of the process variables at any number of designa levels but also illuminates the dependence of each of these effects on values assigned to the remaining variables. Furthermore, the kn ledge of such primary and higher-order effects is characterised maximum precision and reliability.

The utility of a factorial experiment in such a complex prac situation is exemplified in the present study, which was devoted to influence of variations in four process-variables on the efficiency decontamination process for removing radioactive isotopes fro liquid waste. The process employed barium sulphate and carbo principal scavenging agents. The ultimate object of the study wa adjust operating conditions in such a manner that the major por of the activity became associated with the solid phases present. Th could then be filtered off and disposed of by burial.

Description of Waste and Process

The liquid waste employed for the study was a general laborat waste thought to be typical of the material to which the decontami tion process would be applied. The principal radioactivity of the so tion was attributable to salts of radium, thorium, actinium, and th

* Mound Laboratory is operated by Monsanto Chemical Company for the United S Atomic Energy Commission under Contract No. AT-33-1-GEN-53.

aughters. The total count on the waste was 44 640 counts per minute er millilitre, measured as alpha plus beta. Of this total, a figure of 1 230 counts per minute per millilitre was registered as alpha only. The liquid was not clear and contained much suspended matter. The oncentration of inorganic salts was of the order of five parts per million, nd considerable quantities of citrate or tartrate were present. The H of the liquid was 8·3.

The decontamination process which was applied to the waste had lready given satisfactory results with a waste of more specific character. As adapted to the present study, the first step consisted of adding arium chloride solution (100 g of $BaCl_2.2H_2O$ per l.) to 250 ml. of the nfiltered waste until the barium chloride concentration reached a reviously prescribed value. A specified quantity of aluminium reagent 100 g of $Al_2(SO_4)_3.18H_2O$ per l.) was next added, followed imme-iately by a designated quantity of carbon slurry (20 g of carbon per l.). Either N sodium hydroxide or N hydrochloric acid was then added until he *p*H attained the 'final *p*H' value chosen for the experiment. The olid phases present were then allowed to stand in contact with the iquid for ten minutes before filtering. The liquid was agitated con-inuously by magnetic stirring both during the addition of reagents and uring the ten-minute contact period. A one-millilitre sample of the ltrate was mounted on a stainless steel slide and counted for alpha and hen for alpha plus beta, the beta count being obtained by difference. Counting operations were completed within six hours after mounting.

Factors and Factor Levels

The efficiency of the process was defined as the activity remaining, ither alpha or beta, smaller activities being associated with higher fficiencies. The object of the study was to determine the dependence f the efficiency on four process-variables: the amount of barium hloride added, *B*; the amount of aluminium sulphate added, *A*; the mount of carbon added, *C*; and the final *p*H, *P*. Each of these factors as studied at two levels: *B*, 0·4 and 2·5 g of $BaCl_2.2H_2O$ per l.; , 0·4 and 2·5 g of $Al_2(SO_4)_3.18H_2O$ per l.; *C*, 0·08 and 0·4 g of carbon er l. ; and *P*, 6 and 10.

Nature of the Problem

If all physical and chemical constants associated with this process ere known, then one could conceivably develop a general mathe-natical theory relating the efficiency of the process to the value of the our process-variables chosen for study. The theory could be tested in he laboratory by applying the process to a waste of known composi-on. A relatively few tests, executed in accordance with the classical ictum to vary but one factor at a time, might thus have led to an cceptance or modification of the theory and an understanding of the ausal relationships governing the process.

The real situation was otherwise. By no means all of the phys constants were known, nor was the exact composition of the wa available. Finally, even if this information had been at hand, a effort to combine the algebraic relations involved into a comprehen theory would have led to unwieldy expressions of excessive complex Clearly this was a case where a more empirical approach, with f utilisation of the theory of complex experiments, was in order.

Experimental Plan and Results

In order to reveal all interrelated effects as well as to achieve ma mum reliability for the conclusions, a factorial plan was adopted the study. A factorial experiment, by definition, comprises a set of te in which all possible combinations of factor-levels are represented equal number of times. With four factors, each at two levels, $2^4 =$ possible combinations exist.

TABLE I

A 2^4 factorial experiment in four blocks of eight tests with PBAC interaction confounded with blocks

Block I							
$P_1B_1A_2C_2$	$P_2B_2A_2C_2$	$P_1B_2A_1C_2$	$P_1B_1A_1C_1$	$P_1B_2A_2C_1$	$P_2B_2A_1C_1$	$P_2B_1A_1C_2$	$P_2B_1A_$
Block II							
$P_1B_1A_1C_2$	$P_2B_1A_1C_1$	$P_2B_2A_1C_2$	$P_1B_1A_2C_1$	$P_2B_2A_2C_1$	$P_1B_2A_1C_1$	$P_1B_2A_2C_2$	$P_2B_1A_$
Block III							
$P_2B_2A_1C_1$	$P_2B_1A_2C_1$	$P_1B_2A_2C_1$	$P_1B_1A_1C_1$	$P_1B_1A_2C_2$	$P_2B_2A_2C_2$	$P_1B_2A_1C_2$	$P_2B_1A_$
Block IV							
$P_2B_1A_1C_1$	$P_1B_2A_2C_2$	$P_1B_1A_2C_1$	$P_1B_2A_1C_1$	$P_2B_2A_1C_2$	$P_2B_1A_2C_2$	$P_1B_1A_1C_2$	$P_2B_2A_$

The complete plan of the experiment is given in Table I. T factorial experiment was duplicated so that a total of thirty-two t was made. These correspond to the thirty-two cells of the table. T test symbol in each cell employs the four factor symbols with subscri '1' or '2', depending on whether the factor was assigned its lower higher value, respectively.

Only eight tests could be run through and counted in a single d The four blocks of Table I correspond to the four days required complete the experiment. Blocks I and II comprise one compl replicate of the experiment and Blocks III and IV comprise the seco replicate. Blocks I and III are duplicate blocks, and similarly Blocks II and IV. The division of the thirty-two blocks into four blo of eight in the manner shown facilitated reduction of experimental err

The order in which the eight treatment symbols appear in each block orresponds to the actual temporal order in which the runs were arried out. This order was chosen at random. Randomisation reduced ne possibility that bias would be present in the comparisons to be rawn from the data and also ensured a valid estimate of experimental rror.

he Main Effects of the Factors

To formulate the various effects of the experiment, the results were laced in systematic order in Table II. The effects were expressed as inctions of the average counts for duplicate runs.

TABLE II

Results of the 2^4 factorial: Activity remaining
(tens of counts per minute per millilitre of filtrate)

Treatment	Blocks I and II		Blocks III and IV		Average		Per cent of Original Activity Remaining	
	α	β	α	β	α	β	α	β
$P_1B_1A_1C_1$	881	1 437	834	1 236	858	1 337	40	57
$P_1B_1A_1C_2$	650	1 024	494	1 009	572	1 017	27	43
$P_1B_1A_2C_1$	191	574	257	837	224	706	11	30
$P_1B_1A_2C_2$	183	716	193	720	188	718	9	31
$P_1B_2A_1C_1$	289	579	178	563	233	571	11	24
$P_1B_2A_1C_2$	188	498	163	537	175	518	8	22
$P_1B_2A_2C_1$	225	527	370	762	298	645	14	28
$P_1B_2A_2C_2$	135	507	156	577	145	542	7	23
$P_2B_1A_1C_1$	1 180	1 364	1 193	1 528	1 187	1 446	56	62
$P_2B_1A_1C_2$	1 039	1 433	1 146	1 306	1 093	1 370	51	59
$P_2B_1A_2C_1$	466	906	890	1 144	678	1 025	32	44
$P_2B_1A_2C_2$	781	1 130	775	1 078	778	1 104	37	47
$P_2B_2A_1C_1$	298	639	273	512	286	576	13	25
$P_2B_2A_1C_2$	238	475	254	499	246	487	12	21
$P_2B_2A_2C_1$	420	664	429	700	425	682	20	29
$P_2B_2A_2C_2$	350	686	389	660	370	673	17	29

The symmetry of the factorial scheme permits the effect of each ctor to be formulated under each of the $2^3 = 8$ possible combinations values of the remaining three factors. The average of these eight mple effects is called the 'main effect' of the factor. The main effect

so defined is equivalent to the excess of the average count for the ei treatments employing the high level of the factor, over the averag the eight employing the low level. The main effect of each of the f factors, for both alpha and beta activity, is given in the first sectio Table III. A negative value of the effect, provided that it is la enough to be significant, indicates that more activity remained w the lower level of the factor was employed, so that raising the leve the factor was beneficial.

TABLE III

Main effects and interactions of the 2^4 factorial
(Results in tens of counts per minute per millilitre of filtrate)

Effect	Value of Effect α	Value of Effect β
Main effects:		
P	296*	164*
B	− 425*	− 504*
A	− 193*	− 153*
C	− 78†	− 70†
Two-factor interactions:		
PB	− 178*	− 128*
PA	53	55
BA	267*	251*
PC	55	46
BC	1	6
AC	42	65†
Three-factor interactions:		
PBA	4	− 6
PBC	− 27	− 31
PAC	3	− 6
BAC	− 69†	− 57†
Four-factor interaction:		
PBAC	17	38

* Attains 1 per cent significance level (α, 94; β, 76).
† Attains 5 per cent significance level (α, 68; β, 56).

Interactions

The six two-factor interactions, each measuring the dependence the effect of a given factor on the value assigned to a second factor, entered in the second section of Table III. Each of the four th factor interactions, comprising the third section, measures the dep dence of a two-factor interaction on the value of a third factor.

The single four-factor interaction formally corresponds to the dependence of any three-factor interaction on the value of the remaining factor. However, in the experimental design depicted in Table I this interaction has been confused or 'confounded' with the differences between the four blocks or days. The loss of any knowledge of the four-factor interaction is not serious since it is highly unlikely that such a high-order effect could be real. The advantage gained is sensational: every one of the remaining fourteen factorial effects of Table III has been rendered independent of blocks, i.e. day-to-day differences.[1]

The principle of confounding greatly alleviates one of the principal disadvantages of factorial experiments: the possibility that, because of the large number of tests required, experimental conditions may vary and vitiate the comparisons to be made. In the present experiment, an unforeseen change in operators was necessary at the end of the first day's runs. Because of the confounding employed, any resulting change in operating techniques could have no influence on the effects studied by means of the factorial experiment.

Statistical Evaluation of the Effects

The efficiency of the factorial experiment is shown by the fact that all of the data are utilised in the formulation of each main effect and each interaction. Each effect is formulated as the difference between the average activity remaining for eight of the treatments and the average for the remaining eight. Thus the effects are estimated with maximum precision.

If the various differences, which formally express the effects, do not actually correspond to any real causal relation between the efficiency of decontamination and the process-variables, they will vary about zero in a random manner determined by the error present in the experiment. The further a measured contrast is from zero, the less likely it is that random error alone is producing it; or the more likely it is that the corresponding effect is real. The differences between duplicate treatments may be made to furnish a measure of the experimental error from which day-to-day variations have been removed by confounding. This experimental error may be expressed in terms of critical values which the computed effects must equal or exceed in order to be considered statistically significant and probably real. These values are given at the bottom of Table III. Thus for the alpha activity there are but 5 chances in 100 that the experimental error could produce an apparent effect whose absolute value is as large as 68 (5 per cent significance level) and but 1 chance in 100 that the error could give an apparent effect as large in absolute value as 94 (1 per cent significance level).

Discussion of Results

The treatment results exhibited preferential removal of alpha activity. This is evident from the last two columns of Table II, in which the

activity remaining for each treatment is expressed as a percentag
that in the untreated waste. In spite of this preference, Table III sh
that the factorial effects, obtained by contrasting treatments, wer
the same order of magnitude for the two types of activity. Also,
significance levels attained in the two cases, except for the doubtful
interaction, were precisely the same.

Marked overall beneficial effects of raising either the bari
chloride concentration or the aluminium sulphate concentration
evident from the significantly negative main effects of these fact
(B = — 425, — 504; A = — 193, — 153). For each type of activ
the ratio of the main effects is of the same order of magnitude as
ratio of increments in molar concentration for the two salts, corresp
ing to the changes in factor levels:

$$\frac{B_2 - B_1}{A_2 - A_1} = \frac{10{\cdot}23 - 1{\cdot}64}{3{\cdot}76 - 0{\cdot}6} = 2{\cdot}77$$

Important additional information relative to these factors is suppli
by the interaction between them (BA = 267, 251). The interactio
depicted for the alpha activity in Table IV. The fact that this int
action is significantly positive indicates that the beneficial effect
increasing either factor is favoured by operating at the lower level
the other.

This interaction is readily explained by computing the concentrat
of excess barium or sulphate ion present for each of the four alumini
sulphate-barium chloride concentration combinations in Table
The results of the calculation are given in Table VII. Thus the mar
beneficial effect on alpha activity removal of increasing the bari
chloride concentration (— 692), while keeping the aluminium sulph
concentration at its lower level, must be attributed to the excess bari
ion created (8·4 millimoles per l.) since the change in quantity
barium sulphate precipitated (1·6–1·8 millimoles per l.) was q
small. When the aluminium sulphate concentration was high,
beneficial effect of increasing the barium chloride was much sma
(— 158). This is attributed to the marked decrease in excess sulph
ion present (9·6–1·0 millimoles per l.) which was more than suffic
to counterbalance the sixfold increase in barium sulphate formed.

Analysis of the interaction in terms of the dependence of the ef
of aluminium sulphate concentration on the value assigned to
barium chloride leads to similar conclusions. The beneficial eff
(— 460) of increasing the aluminium sulphate concentration wh
maintaining the barium chloride at its lower level was solely the res
of the increase in excess sulphate ion concentration (0·2–9·6 millimo
per l.) since the quantity of barium sulphate precipitated remai
unchanged (1·6 millimoles per l.). Likewise, any beneficial effect
the sixfold increase in barium sulphate (1·8–10·2 millimoles per
accompanying the increase in aluminium sulphate concentration at

TABLE IV

BA two-way table for alpha count

(*Results in tens of counts per minute per millilitre of filtrate*)

	$B = B_1$	$B = B_2$	$B_2 - B_1$
$A = A_1$	927	235	− 692
$A = A_2$	467	309	− 158
$A_2 - A_1$	− 460	74	

$BA = \frac{1}{2}[-158 - (-692)] = \frac{1}{2}[74 - (-460)] = 267.$

TABLE V

PB two-way table for alpha count

(*Results in tens of counts per minute per millilitre of filtrate*)

	$B = B_1$	$B = B_2$	$B_2 - B_1$
$P = P_1$	460	213	− 247
$P = P_2$	934	331	− 603
$P_2 - P_1$	474	118	

$PB = \frac{1}{2}[(-603 - (-247)] = \frac{1}{2}[118 - 474] = -178.$

TABLE VI

BAC three-way table for alpha count

(*Results in tens of counts per minute per millilitre of filtrate*)

		B		
C	A	B_1	B_2	$B_2 - B_1$
C_1	A_1	1 022	260	− 762
	A_2	451	361	− 90
C_2	A_1	832	211	− 621
	A_2	483	258	− 225

$BAC = \frac{1}{4}\{[-225 - (-621)] - [-90 - (-762)]\} = -69.$

TABLE VII

Quantity of $BaSO_4$ formed and concentration of excess ion for various barium chloride–aluminium sulphate concentration combinations

(*$BaSO_4$ and excess ion expressed in millimoles per litre*)

	$B = B_1$	$B = B_2$
$A = A_1$	1·6 $BaSO_4$ 0·2 SO_4^{--}	1·8 $BaSO_4$ 8·4 Ba^{++}
$A = A_2$	1·6 $BaSO_4$ 9·6 SO_4^{--}	10·2 $BaSO_4$ 1·0 SO_4^{--}

higher barium chloride concentration was completely neutralised the reduction in concentration of the excess ion (8·4 Ba^{++} to 1·0 SO_4

Since similar results were obtained with the beta activity, one c conclude that over the range of conditions embraced by the exp ments, the amount of barium sulphate formed is of secondary imp tance to the amount of excess ion present. Presumably, the bari sulphate, by sorption of the excess ion, assumes the charge of the la and thereby becomes effective in removing oppositely charged from solution. Since both excess barium ion and excess sulphate are beneficial, we may conclude that both radioactive anions radioactive cations were present in the original waste.

The BA interaction cannot be attributed to changes in the concen tion of aluminium or chloride ion. This is because the increase chloride ion concentration, accompanying the increase in bari chloride concentration, is independent of the aluminium sulphate c centration. Conversely, the increase in aluminium ion concentrati accompanying the increase in aluminium sulphate concentration independent of the barium chloride concentration. In any event, aluminium and chloride ions would be expected to play only a mi role as potential determining ions because of their lack of similarity the lattice ions of the precipitate.[2]

The significant *p*H—barium chloride concentration interact (PB = — 178, — 128), illustrated for the alpha activity in Table shows that the beneficial effect of increasing the barium chloride fa was accentuated by increasing the *p*H. Again referring to Table V the increase in barium chloride concentration is seen to be acc panied, on the average, by substitution of excess barium ion for ex sulphate ion, that is, by conversion of negatively charged bari sulphate to positively charged barium sulphate. Any improvemen decontamination resulting from this change should be accentuated increasing the *p*H from 6 to 10, since such an increase would favour formation of anions of amphoteric radio-isotopes such as lead, bism and polonium.

Although, so far as the *p*H—barium chloride concentration in action is concerned, increasing the *p*H from 6 to 10 was beneficial, significantly positive main effect of this factor (P = 296, 164) sh that the overall effect of the increase was detrimental. Other fac were evidently sufficient to counteract the beneficial PB interacti The mechanisms operating here are suggested by the positive *p*H aluminium sulphate concentration interaction (PA = 53, 55) and positive *p*H—carbon concentration interaction (PC = 55, 46). Both these nearly attained the five per cent significance level. The posit PA interaction is doubtless related to the action of an auxiliary scave ing agent, aluminium hydroxide. More of this substance sho precipitate at *p*H = 6 than at *p*H = 10 because of the amphote nature of aluminium. The positive PC interaction suggests that carbon functions more effectively at the lower *p*H.

The overall effect of increasing the carbon concentration, though naller than that of aluminium sulphate or barium chloride, was efinitely beneficial (C = — 78, — 70). It is possible that the carbon ffect might have been relatively greater had this reagent been added efore instead of after the barium sulphate was precipitated.

The negative barium chloride-aluminium sulphate—*p*H interaction BAC = — 69, — 57) is formulated for the alpha activity as a three-ay table in Table VI. This table shows that the undesirable positive A interaction is reduced by operating at the higher carbon concen-ation.

onclusions

The factorial experiment has established certain relationships be-veen the extent of decontamination and the four process-variables in highly efficient manner. By exploring interactions between the ctors, mechanisms by which the factors operate have been suggested, us leading to an improved understanding of the process.

cknowledgements

The authors gratefully acknowledge the fundamental work of Mr M. Hamilton, Mr Richard Deem, and Mr James Poppleton.

REFERENCES

Cochran, W. G. and Cox, G. M. (1950). *Experimental Designs*, pp. 159–64. Wiley, New York.

Wahl, A. C. and Bonner, N. A., Editors (1951). *Radioactivity applied to Chemistry*, p. 125. Wiley, New York.

AN EXTENSION OF THE USE OF THE χ^2-TEST

ERIK K. LEANDER and DAVID J. FINNEY
Department of Statistics, University of Aberdeen

A recent inquiry about the statistical analysis appropriate
problem in paper manufacture led to a recognition of a si
extension of the use of χ^2 in the analysis of sample statistics
Poisson distributions. Since this extension may not be gene
known, the authors have in this article described the test
reference to the actual problem.

The Experiment

Experiments have been performed for comparing alternative meth
of removing dirt contained in the suspension of fibre from which pa
is to be made. The experimental technique is to contaminate the
pension deliberately, and then to make sheets of paper from this
pension before and after it has been cleaned. Dirt spots are cou
by eye, and the relation between the number of spots per sheet f
the cleaned and uncleaned material is considered to be an estima
the effectiveness of the cleaning.

In a typical experiment, two cleaners are compared. It w
obviously be desirable to have both tried on portions of the same
taminated suspension, but for technical reasons this is not gene
possible. Although the experimenter will attempt to make the in
contamination approximately the same in both suspensions, the sta
cal analysis must take account of any inequality, and the theory
follows is quite general in this respect.

In consequence, the comparison will involve four samples of pa
one from dirty and one from cleaned suspension for each cleaner.
numbers of sheets of paper in the four samples can be chosen by
experimenter entirely independently of one another. The general f
of experimental result can be represented by

	I		II	
	Number of Sheets Made		Total Number of Di Spots	
	Before Cleaning	After Cleaning	Before Cleaning	After Cleanir
Cleaner A	m_A	n_A	x_A	y_A
Cleaner B	m_B	n_B	x_B	y_B

itistical Analysis

Let the expected densities of dirt spots per sheet be t_A, t_B before aning and pt_A, pt_B after cleaning; this corresponds to a null hypoesis of equal effectiveness, p being the proportion of dirt remaining er the cleaning process. Assuming the counts per sheet to have isson distributions in each of the four cases, we obtain the following pected values and probability densities for the total dirt counts (subipts omitted):

Expected Count		Probability Density	
Before Cleaning	After Cleaning	Before Cleaning	After Cleaning
tm	ptn	$\frac{(tm)^x}{x!}e^{-tm}$	$\frac{(ptn)^y}{y!}e^{-ptn}$

The marginal totals $(x_A + y_A)$ and $(x_B + y_B)$ contain no information p, since t_A, t_B are unknown. Hence, for the purposes of the test of nificance, we may consider these totals as constants, k_A and k_B, actly as in an ordinary 2×2 contingency table. For either cleaner conditional probability of the appropriate (x, y) can then be itten, apart from insertion of the A or B suffix throughout,

$$\frac{\frac{(tm)^x(ptn)^{k-x}}{x!(k-x)!}e^{-(tm+ptn)}}{\frac{(tm)^x(ptn)^{k-x}}{x!(k-x)!}e^{-(tm+ptn)}} = \frac{\frac{k!}{x!(k-x)!}(tm)^x(ptn)^{k-x}}{(tm+ptn)^k}$$

$$= \frac{k!}{x!(k-x)!}\left[\frac{m}{m+pn}\right]^x\left[1-\frac{m}{m+pn}\right]^{k-x}$$

d thus the conditional distributions of x_A, x_B are binomial. Furtherre, if

$$\frac{m_A}{n_A} = \frac{m_B}{n_B}, \quad \text{then} \quad \frac{m_A}{m_A + pn_A} = \frac{m_B}{m_B + pn_B}$$

nce, if the proportions between the number of sheets made before d after cleaning are equal for both cleaners, our null hypothesis erts that two binomially distributed variables have the same probaty parameter. We may treat II above as an ordinary contingency le, and—employing Yates's correction—we obtain the usual test istic:

$$\chi^2 = [|h| - 0{\cdot}5]^2\left[\frac{1}{x_A - h} + \frac{1}{y_A + h} + \frac{1}{x_B + h} + \frac{1}{y_B - h}\right]$$

ere $h = x_A - \frac{(x_A + y_A)(x_A + x_B)}{x_A + y_A + x_B + y_B}$ and χ^2 has 1 degree of freedom.

There are obvious possibilities of extensions to cases of $r \times 2$ and $r \times c$ contingency tables.

Because of the simplicity of the test, the experimenter should mally choose his number of sheets so that $m_A/n_A = m_B/n_B$. How an inequality may arise on account of the loss of a sheet of pap simply because exact equality would imply an undesirable si experiment. Even then a test on the same principle should be factory, but a slightly more involved computation of the maxi likelihood estimate of $\hat{p}$, the proportion of dirt remaining (unde null hypothesis of equal effectiveness of cleaners), will be requ This estimate is the positive root of the quadratic

$$n_A n_B(x_A + x_B)\hat{p}^2 + [n_A m_B(x_A - y_B) + n_B m_A(x_B - y_A]\hat{p} - m_A m_B(y_A + y_B)$$

With this value of $\hat{p}$:

$$h = \frac{n_A \hat{p} x_A - m_A y_A}{m_A + n_A \hat{p}}$$

and χ^2 is calculated as before. As a check, it may be noted tha arrive at the same test-statistic as previously if $m_A/n_A = m_B/n_B$.

Numerical Example

No data from experiments of the type described are availabl analysis. The following hypothetical figures, bearing no relatio the particular experimental situation, will be used to illustrat computations:

	I Number of Sheets		II Number of Spots	
	Before Cleaning	After Cleaning	Before Cleaning	After Cleani
Cleaner A ..	10	33	340	80
Cleaner B ..	12	40	360	120

Estimates of p for the two cleaners separately are obviously

$$\hat{p}_A = \frac{10 \times 80}{33 \times 340} = 0{\cdot}071\,3$$

$$\hat{p}_B = \frac{12 \times 120}{40 \times 360} = 0{\cdot}100\,0$$

The difference between these appears to be quite substantial, and χ^2-test is needed. The quadratic equation is easily formed as

$$924\,000\hat{p}^2 + 199\,120\hat{p} - 24\,000 = 0$$

of which the only solution between 0 and 1 is $\hat{p} = 0{\cdot}086\,1$.

mula in the previous section then gives $h = 166{\cdot}0/12{\cdot}84 = 12{\cdot}93$, ence can be constructed the expected frequencies:

	Before	After
A	327·07	92·93
B	372·93	107·07

ese have the same marginal totals as before, and the correctness of easily verified by checking that each line taken in association with appropriate m, n corresponds to $\hat{p} = 0{\cdot}086\,1$. Hence

$$\chi^2 = (12{\cdot}93 - 0{\cdot}5)^2\left(\frac{1}{327{\cdot}07} + \frac{1}{92{\cdot}93} + \frac{1}{372{\cdot}93} + \frac{1}{107{\cdot}07}\right)$$

$$= 0{\cdot}472 + 1{\cdot}663 + 0{\cdot}414 + 1{\cdot}443$$

$$= 3{\cdot}99$$

e difference between values of $\hat{p}$ for the two cleaners is therefore istically significant at the 5 per cent probability level.

erimental Design

he cost of the experiment on paper cleaners is largely determined he total number of sheets to be made for test purposes. Hence it nportant to try to maximise the power of the test for a fixed number neets. Normally, the pre-experimental information will allow for a onably good guess at the proportion of dirt spots removed, and n information will provide an indication of how efforts should be ributed between sheets from uncleaned and from cleaned suspen-s. If the experimenter knows little or nothing beforehand about the tive merits of the cleaners being tested, he will not discriminate veen them in the design of the experiment. The numbers of sheets aper will then be chosen so that $m_A = m_B$ and $n_A = n_B$, and an rt will be made to obtain equal initial contamination in the two ensions. The problem is to choose

$$\xi = \frac{n_A}{m_A + n_A} = \frac{n_B}{m_B + n_B}$$

ect to the condition that the total number of sheets is fixed, so as btain a high power for the test.

llitto[1] has indicated an approximation to the power function of χ^2-test of equality in probability parameters for two binomially ibuted variables. This approximation may be employed to obtain iterion of the optimum proportion of sheets from uncleaned and

from cleaned suspension. The power will be approximately maxi
for the value of ξ which minimises the factor

$$\frac{1}{m} + \frac{1}{\frac{1}{2}(p_A + p_B)n} = \frac{1 - (1 - p_0)\xi}{p_0(1 - \xi)\xi},$$

where $p_0 = \frac{1}{2}(p_A + p_B)$ is the average proportion, for the two cle
of dirt remaining after cleaning. The minimum will be attaine

$$\xi = \frac{1}{1 + \sqrt{p_0}},$$

so that a guess at the average proportion of dirt remaining will ind
in what proportion sheets should be made in order to obtain an effi
design.

The condition on ξ is rather crudely obtained. We are maxim
the absolute value of Sillitto's μ as a function of the expected va
of total dirt counts from sheets made before and after cleaning.
means that we are not maximising the expected power but the p
obtained for expected values. It also involves ignoring the probab
of rejecting the null hypothesis by obtaining a significant valu
$\hat{p}_A - \hat{p}_B$ having the wrong sign. For the purpose at hand, however
accuracy should be quite sufficient. A cruder approach, leading t
same conclusion, is to note that for each suspension my/nx is an esti
of p; an approximate variance for the difference of the estimate
easily be derived, and this is minimised by the same choice of ξ.

We wish to extend our thanks to the Wiggins Teape Group Res
Organisation for their kind permission for us to refer to the experi
on paper cleaners carried out by the Organisation.

REFERENCE

[1] Sillitto, G. P. (1949). 'Note on approximations to the power function
"2 × 2 comparative trial"', *Biometrika*, **36**, 347.

QUESTION AND ANSWER

Readers are invited to submit questions, or to deal shortly with points of general interest by supplying both the question and the answer.

stion

The performance of an aeroplane depends on the air temperature, and t is necessary to study the climatological characteristics of a particular port in order to determine the limitations, if any, on the full exploitation he possibilities of the aircraft. For instance, if it is desired to estimate take-off weight that may be exceeded in summer on 20 per cent of asions, it is necessary to know the eighteth percentile of the distribution emperature in the summer and at the usual time of departure. I was s led to the following statistical problem:

To estimate the percentiles of the distribution of temperatures at airport t, say, 6 p.m. in July.

The difficulty is that the observations on successive days are not indepen- t; several tests of mine showed that there was quite a marked degree of ocorrelation.

f the daily records of the local meteorological office are available, the owing procedure will, I think, be satisfactory. Select a moderately large mple in such a way that the individual observations may be considered ndependent (e.g. three days at ten-day intervals from each July month he last twenty years, the first day in each month being chosen at random) treat the resulting data by the usual methods (the population may be umed to be normal).

But usually the data are more scarce—e.g. a grand mean based on all y observations made in July at 6 p.m. during several years, and a ndard deviation which I view with suspicion because I do not know how as been calculated; I strongly suspect that in most cases it is the standard iation of the distribution of the monthly means. Obviously this should multiplied by $\sqrt{31}$ if the observations were independent, but since they not, this introduces another difficulty. Sometimes not even as little as is given—just a grand average of daily means and averages of daily xima and daily minima.

t remains to be said that I would be content with point estimates of the centiles and would readily dispense with their standard errors.

wer

f daily records are available, the course of action you suggest is valid, ugh there is nothing to be gained by having any interval between obser- ons and thus discarding good data. It would be perfectly justifiable to struct a cumulative percentage frequency graph for all the daily ervations of temperature near airport X at 6 p.m. in July which are ilable, and to read the percentiles directly from this graph; or, as you

suggest, to fit a frequency curve to the data and deduce the perce from that. The autocorrelation does not affect the validity of such estim but merely makes their sampling error as estimates greater than it w have been if the daily observations used had varied randomly; an sampling error will certainly be smaller than it would be if you le some of your data in order to secure approximate randomness in the

If, however, you are given only a grand mean for July at 6 p.m. an standard deviation $\sigma_{\bar{x}}$ of means for July at 6 p.m. in different years, will have to introduce some more information to give an answer. Sin

$$\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{31}}\sqrt{\left[\sum_{s=-30}^{+30}\left(1-\frac{|s|}{31}\right)\bar{\rho}_s\right]}$$

where σ is the standard deviation of daily figures for July at 6 p.m. a the average correlation between daily figures in July separated b interval of s days, you will have to supply an approximate value for the f on the right-hand side in order to estimate the unknown σ from the g $\sigma_{\bar{x}}$. It may be possible to infer this from experience of other compar stations where the existence of daily records makes it possible to investi $\bar{\rho}_s$ for different values of s, fit a curve to it, and evaluate the factor. T having obtained an estimate of σ, you could use it, in conjunction wit grand mean given, to fit your normal curve and deduce the percentile

In the last state of affairs you describe your information is very inadeq To produce an answer you would not only have to assume a value based on your experience of comparable stations, but you would ha assume a form for the diurnal cycle to enable you to adjust the given to 6 p.m. before you could produce a normal curve from which to de your percentiles.

G. H. JO

INDUSTRIAL APPLICATIONS SECTION OF THE ROYAL STATISTICAL SOCIETY

List and Reports of Meetings

BIRMINGHAM AND DISTRICT GROUP

January 1956. Seasonal Variation in Time Series: J. R. N. Stone

February 1956. Problems of Optimum Production Run: T. E. Easterfield

GLASGOW GROUP

h January 1956. Simplified Statistics in Industry: J. S. Ogilvie

d February 1956. Control of Stocks: J. Hepburn

LEICESTER GROUP

d January 1956. Criminology: L. T. Wilkins

LONDON GROUP

February 1956. Shortened Life Testing: G. A. Barnard

l March 1956. The Handling of Uncontrolled Data: G. H. Jowett

MERSEYSIDE GROUP

January 1956. Ideas for Managers: Probability and Policy: M. J. Moroney

February 1956. Some Problems in Consumer Testing of Convenience Goods: N. J. Squirrell

NORTH-EASTERN GROUP

December 1955. Elementary Principles of Survey Design: G. T. Park

h January 1956. Some Aspects of Personnel Selection: W. Sluckin

February 1956. Statistical Method in some Hospital Administration Problems: D. J. Newell

SHEFFIELD GROUP

d February 1956. Stochastic Processes: D. A. Hopkins

h March 1956. The Theory of Games: D. V. Lindley

SOUTH WALES GROUP

December 1955. Inventory Management: K. S. Lomax

This lecture was substantially the same as that given to the Sheffield oup on 27th January 1955 and reported in *Applied Statistics*, **4**, 124.

h January 1956. Twins and the Study of Heredity: J. A. H. Waterhouse

d February 1956. The Analysis of Continuous Measurements: K. W. Kemp

Dr Kemp dealt first with the problems that arise in the statistical analysis continuous records. He described his own search for a speedy method making such analyses and explained the principles of a simple instrument ch he showed to be capable of yielding a qualitative as well as a quantita- picture of the types of variation under examination. The results were ained with far greater speed and reliability than is possible using correlo- m methods.

STUDY SECTION OF THE ROYAL STATISTICAL SOCIETY

Lists and Reports of Meetings

LONDON GROUP

14th December 1955. Labour Wastage in Industry: H. Silcock

11th January 1956. Local Population Projections in England and W
J. R. L. Schneider

8th February 1956. Recent Market Research Methodology: D. L. Lamb

BRISTOL GROUP

15th December 1955. Statistics in Hospital Planning and Design: N.
Bailey

It is hoped that a paper on this subject will appear in a later issu
Applied Statistics.

19th January 1956. The Dual Problem in Linear Programming: H
Mackenzie

Linear programming consists of (i) a mathematical formulation of
nomic problems, and (ii) the numerical solution of those problems.
solution can provide not only the optimum plan in a given situation
also a means of assessing the cost of departure from the optimum, a
quick method of finding new optima when the economic situation cha
Two methods of formulating the problem—the direct (linear programr
method and the dual method—were illustrated by means of a simple
duction-scheduling example, and it was shown that, in application
method of formulation chosen should be related to problems likely to ar
the future. Applications of the dual discussed were: alternative progran
valuation of assets, subcontracting, and decentralisation problems. A
discussed was how one can formulate the dual problem when the l
programming problem contains equations instead of inequalities. A po
answer is to express each equation as *two* inequalities. Thus the equ
$x_1 + x_2 = 1$ could be written: $x_1 + x_2 \leqslant 1$, $-x_1 - x_2 \leqslant -1$ in a r
misation problem. The dual problem can then be set up in the normal
by matrix transposition.

16th February 1956. Total and Difference in Samples of Two: M. J. Mor

Mr Moroney discussed the logical and practical processes involve
using statistical inferences from data, illustrating with examples from qu
control and industry in general. He stressed the proper relationshi
theory and practice in statistical work.

BOOK REVIEWS

. Bush and F. Mosteller: *Stochastic Models for Learning*. New York: John Wiley and Sons; ndon: Chapman and Hall, 1955. Pp. xvi + 365. 72s.

'his book gives an account of the authors' work on the development of eoretical framework for analysing psychological experiments in learning he simpler variety) with animal or human subjects. While of considerable rest to the mathematical statistician as a guide to the kind of data ilable and to the new problems of analysis they create, the book seems nded even more for those experimental psychologists prepared to study athematical and statistical monograph of this technical level (which, in e of the authors' remarks in the preface, is a rather formidable one). psychologists' task has been slightly helped by starring the more theo- cal or recondite sections; the book is, moreover, divided into two parts, theory and applications respectively, of which the second part is the e relevant for psychologists.

'he amount of detailed information, both theoretical and applied, is siderable, and it is not possible to summarise it all in a short review. It , however, be explained that one of the theoretical complexities in dels for learning is that the probabilities of the possible responses are ditioned by events; e.g. by an electric shock following a wrong response y food following a correct one. Consequently models for learning are particularly simple stochastic processes, for even if the learning process ssumed Markovian in the (continuous) state variable *characterised by the nse probability p itself*, this is a new type of process which the authors find ore convenient to study in terms of conventional Markov chains whose crete) states are the possible outcomes of any trial, but whose transition babilities are continually being modified by the 'conditioning' referred Such modifications are handled theoretically by means of operators ich are for simplicity assumed linear). Part I develops this theory some- t fully before applying it in any detail to the data discussed in Part II, an which has perhaps some disadvantages as well as advantages as it ls to discourage a complete reading of Part I. References are given at end of each chapter, and the book is attractively printed.

he methods developed in this monograph represent a unified and com- ensive approach to the analysis of learning experiments. As, however, are based on purely empirical models, they could perhaps be criticised noring other crucial aspects of the theory of learning. Future develop- ts will no doubt tie up with physiological studies of the central nervous m: such studies are already well advanced, and seem likely ultimately e the basis on which more complete models of learning will be con- cted.

M. S. BARTLETT

ge Katona, Lawrence R. Klein, John B. Lansing and James N. Morgan. *Contributions of vey Methods to Economics*. Columbia University Press (London: Geoffrey Cumber- e), 1954. Pp. viii + 269. 40s.

he scope and objects of this volume are rather narrower than its title ld seem to imply. The survey methods referred to are those used in a

single series of annual surveys—the Surveys of Consumer Finances ca
out in the United States by the Survey Research Center of the Univ
of Michigan on behalf of the Board of Governors of the Federal R
System. The methods and results of the surveys are described in the *
Reserve Bulletin, and they have been much used in assessing the c
economic situation in the United States. The present volume goes b
this to consider their use for the purpose of economic research, describ
Klein in his introduction as 'the testing of hypotheses about econom
havior suggested by theory, a knowledge of our institutions, indepe
empirical studies, and other sources'.

Such an empirical testing of hypotheses requires first of all that
should be a reconciliation of the concepts used by the theorists and
used by the providers of the data which are used in the tests. When
data are survey data, the process has to be carried a stage further, an
survey organisers have to devise ways in which their own concepts ca
translated into the ordinary everyday language of the group which is b
surveyed. The survey method will facilitate the reconciliation of theore
and practical concepts, as compared, for instance, with attempts to a
existing collections of data; but existing institutional and other fa
(which may be ignored by the theorist) will have to be taken into acc
in planning the collection of the material. These problems are discu
in the first chapter, by John B. Lansing.

The reconciliation of concepts need not proceed only from the end o
collection of the data. It is possible to reformulate the theory so as to re
it more readily susceptible to testing and to take explicit account of ou
factors—institutional, social, psychological—which are ignored in its 'p
form. This is the subject matter of the second chapter, in which G
Katona discusses a theory of consumer behaviour which takes accou
attitudes and expectations as well as the more obviously economic vari

The next two chapters, both by James N. Morgan, begin the proc
seeking for valid relationships among the data, using theory to suggest
these relationships might be. The techniques used are those of the an
of variance, which can show the existence of relationships without g
them a precise formulation. The author not only describes his method
findings in some detail but carefully considers the question whether his
fulfil the conditions required for the application of variance analysis.

The final stage, the statement of the relationships between variab
the form of a precise model and the estimation of its parameters, is desc
by Lawrence R. Klein. His multi-variate relations use demographic
ables (size of household and age of its head) as well as economic ones.
far the relationships would continue to hold in appreciably different
nomic conditions is a matter of some doubt, and Klein himself suggests
we ought not to expect too much of them. The interest of his two cha
would seem to lie at least as much in their links with what has gone be
and his suggestions for future developments, as in the actual quantit
estimates.

Indeed, the links between all the different sections of the book ar
many ways its most important lesson. Based on long experience, the Su
of Consumer Finances seems to have achieved the integration of the
survey design, and statistical and economic analysis of results, w
is basic to the success of any such piece of work. Its contribution

nomics, as such, must still be considered rather limited; its lessons for conduct of economic or any other kind of research, using survey methods, fundamental.

J. E. G. UTTING

auter and J. S. Downham: *The Communication of Ideas*. Published for the Readers' gest Association. London: Chatto and Windus, 1954. Pp. xviii + 324. 25s.

he book reports in clear English and by means of informative tables a ple survey of Derby, covering reading, radio-listening, club-membership, other leisure activities. It is a descriptive factual study based on an quate and efficient sample in the main, and drawing also upon other ces of data, both official and private. A number of interesting new facts noted and quantified for the first time, but the treatment of these seems nt. The statistical reader will be disappointed in the number of opportties for analysis which have been missed. It could so easily have been uch better book.

he authors show, although the evidence is not very clear, that the major ion of any type of activity is accountable to a small proportion of the ulation, and they attempt to summarise this factor by means of an 'index ommunication'. This was a good idea, but the index has been so overplified that it might be misleading. Ten items which represent comnication potential—club-membership, church-going, pub-going, cinema-g, etc.—are taken, and the sample divided into those who do and those do not go to or participate in the things mentioned. The authors then let those who do not go or participate score zero for such activities. se who do participate they divide into categories of frequency and give points to those who fall into the most frequent category. Thus they de only the non-zero (positive) half of each distribution into deciles, then weight according to the decile. This procedure is odd. It has the ct of weighting according to whether the distribution of frequency is tively or negatively skewed, and the frequency of zero for each activity. account is taken of covariance. The scores for each of the ten activities added. This may not be a serious point in practice when the unequal tment of zero is considered. The point of covariance is perhaps worth mining.

all persons who go to church also go to the cinema ($r_{ab} = +1{\cdot}0$), we shall gain no information by knowing both facts. Similarly where 0, but is some quantity of significance, whether positive or negative in , allowance should be made for the 'overlap'. We might, for example, pose that few people who go regularly to public houses also go regularly hurch (negative $r \neq 0$), then we might consider that those who do go oth have a more varied exposure to communication activities than cted. It would have been interesting to have some analyses of this kind. ould also have been interesting to know something about 'saturation rs'. The authors indicated that a large proportion of each activity accountable to a small proportion of persons, but they did not go on to truct the family of curves which might have presented different charactics for different activities. A simple curve might have been plotted with activity from 1 to 100 per cent as the ordinate and the proportion of persons accounting for the activity as the abscissa. Even to have deed the 50 per cent point for different activities might have been extremely able. Would church-going have the same index as cinema-going? That

is, does cinema-going at the 50 per cent point rely upon a lower prop of its patrons than does church-going for the same point? Such a would have been sociologically valuable—colloquially this would hav us whether cinema-going or church-going was more of a habit with who go.

Here was a case where a few slightly more sophisticated statistica cedures might have been usefully employed, but for some reason the a have been all too cautious and tried to be too straightforward. The was complex, and the simple procedures seem to have left unnoted that had a high probability of being interesting and useful. L. T. W

W. G. Ireson and E. L. Grant (Editors): *Handbook of Industrial Engineering and Manag* Englewood Cliffs: Prentice-Hall (London Agents: Bailey Bros. and Swinfen), 1955. + 1203. 128s.

This book covers a wide field; some idea of its scope may be g from a list of the section headings: Structure of Business Organisat Managerial Economics; Engineering Economy; Manpower Manage and Employment Relations; Motion and Time Study; Factory System Procedures; Industrial Climatology; Factory Planning and Materials H ling; Industrial Standardisation; Tool Engineering; Industrial Sa Industrial Hygiene; Industrial Statistics; Inspection and Quality Co Industrial Operations Research; Industrial Budgeting; Trade Union Industrial Engineering.

The result is a formidable volume. It would be inappropriate i journal to consider all the sections at length, so I propose to conce on the sections with some statistical interest, merely pausing to remar there is no mention of the work done in this country on the statistical a of time study. The section on industrial statistics, occupying nearl pages, is by A. H. Bowker and G. J. Lieberman. It is a very useful e tion of the techniques likely to be of interest to industrial engineers manages to condense a good deal of information into the space avai A wide variety is presented, including control charts, sampling insp schemes, common tests of significance and their operating character correlation and regression, and analysis of variance. Tables and char included, so that the techniques can be applied without reference to sources. There is some overlapping with the section on inspectio quality control, but the latter section, contributed by W. H. Lewis, i cerned to a greater extent with the organisation of quality control with its statistical basis.

The author of the section on operations research is A. W. Swan. He first with its principles and organisation and then presents eight histories, contributed by workers in various fields. The case-histories selected to illustrate the various techniques that may be required; four from Britain and four from the United States. Statistical methods are n in evidence here, in fact the section gives quite a good sample surve one of the major fields in which statistical methods may be applied.

Although the price is high, it is not unduly so for a book of this le It would seem to be a useful reference work for those concerned with fa management in its broadest aspects, and should at any rate find a pla industrial libraries. D. G. B

FOREWORD

E form in which a new journal appears and any success that it may ieve are dependent in large measure upon its first editor. The uncil of the Royal Statistical Society was more than fortunate in uring Mr L. H. C. Tippett as the first editor of *Applied Statistics*, a ition which he has recently relinquished towards the close of the n year of the journal's existence.

Mr Tippett was closely concerned with the creation of the journal, l contributed much to its initial format and character. His experience n author and his wide knowledge of statistics made him particularly l equipped for the editorial work, the excellence of which may be ged by the reader himself. He has been scrupulously careful in serving clarity of style, in avoiding jargon, and in making the sentation as attractive as possible. All this has been achieved with- lowering any statistical standards.

t gives me great pleasure to record on behalf of the Council, and he Society itself, this appreciation of Mr Tippett's valuable years ervice as Editor.

E. S. PEARSON

President

Royal Statistical Society

STATISTICS IN HOSPITAL PLANNING AND DESIGN*

NORMAN T. J. BAILEY

Design and Analysis of Scientific Experiment, 6 Keble Road, Oxford

In this article Mr Bailey shows how statistical methods help to ensure that hospitals make the best use of lim resources. He deals with the calculation of the size of hos required, the achievement of efficiency in hospital design, the use of current statistical returns to check the performa of existing hospitals.

Introduction

The provision of hospital care has a long and chequered hist going back to ancient Greece and Rome in the West and to the ag King Asoka in the East. In the past most of this care appears, fortunately, to have been rather ineffectual. Medical theory and p tice have, however, advanced with increasing rapidity since Renaissance, and at the present time an appreciable part of the sub can claim the status of science. We are therefore now in a positio use modern methods of collective investigation and control to en that this knowledge is used as efficiently as possible. Particul relevant are the methods of operational research developed lar during and since World War II, and the statistical applications cussed in this paper occur largely in the context of this new discip

Although many contributions have been made in recent times to building of better hospitals, it is only in the last few years that a re constructive and integrated approach has been at all possible. By l the Nuffield Foundation and the University of Bristol were alre co-operating in launching such an investigation specifically to enq into the function and design of hospitals. A report[1] of the work has appeared, and represents a major advance in this field. We shall fine ourselves here to aspects that have been found susceptible statistical and mathematical techniques, without embarking on tailed theoretical discussions.

There is, of course, practically no limit in principle to the exten which *ad hoc* improvements can be made in most hospital facilities, in practice medical skill, drugs, building accommodation, finance, are available only to a limited extent. So it is necessary to discover h to reap the best advantage from restricted resources. In a gen

* This article is based on a paper read to the Study Section of the Royal Statistical So (Bristol Group) on 15th December 1955.

cussion like the present one it is convenient to adopt a logical order reatment. We shall therefore consider first methods of assessing how ch medical care is to be supplied to a given area; then go on to look some of the problems involved in trying to make that supply as cient as possible; and finally examine the question of the administra- e use of current statistical returns.

essing the Demand for Medical Care

Inpatients. Suppose we take inpatient care first, and start by thinking w to estimate the demand for it. By 'demand' we mean current uests as opposed to the real *need* for medical care. The latter is a y important problem but has so far received little attention, no ubt because of the great difficulty involved. What actually happens rather complicated. There are several hospitals of different sizes ttered about the country. Patients requiring admission in any area y have a certain choice where to go—depending on their own wishes d the policy of their GPs—but they will in fact tend to go to hospitals t are more rather than less accessible. Conversely, a given hospital geographically compact group of hospitals may receive practically) per cent of all patients requiring admission in the immediate inity. As we go farther away this percentage will fall, until it reaches o in the neighbourhood of the next hospital. Here then is the first difficulty. One is likely to be confronted, not by a neatly delimited a, within which all the people form a closed population at risk though this may happen), but by several overlapping and inter- netrating groups. This picture is a little oversimplified but represents stantially what was found in the Norfolk and Norwich region, for mple. If we use pooled data for a very large area then these edge- ects become relatively less important. On the other hand, the neral character of the area then becomes more heterogeneous and the al estimates of relative demand are correspondingly less useful for diction.

Estimating the Population at Risk. To estimate the population risk for some chosen group of hospitals a convenient procedure is as ows. From lists of addresses of patients admitted to this group ring say the past year, one chooses a convenient sample. Similar mples are then taken from all the hospitals on the periphery of the a that are likely to exert any influence in drawing off patients. us for each district in the area, e.g. urban and rural districts, we can m an estimate of the total demand for medical care represented by, , a patients who went to the main group and b who went to some er hospital. We then use the ratio $a/(a + b)$ as an estimate of the portion of *patients* in that district who are normally hospitalised in ir local hospitals. If the population of the district is N, obtained m census data, then the *effective* population at risk for the main pital can be taken to be $aN/(a + b)$. Summing this quantity over

all districts, both inside and outside the area, gives the estimate, M, the total effective population at risk, e.g.

$$M = \sum \frac{aN}{a+b} \qquad \ldots$$

Suppose the numbers a and b are based on 100 per cent samp addresses. Then, using the fact that a and b are likely to be Po variables, we find that the large sample variance of M is given by

$$\text{var}\,(M) = \sum \frac{abN^2}{(a+b)^3} \qquad \ldots$$

where we are assuming that each N is known nearly exactly. But and b are themselves estimates based on a constant sampling fract λ, of the actual data available, the corresponding expression for variance is, of course,

$$\text{var}\,(M) = \sum \frac{abN^2}{\lambda(a+b)^3} \qquad \ldots$$

Variable sampling fractions give rise to a much more complica formula.

In the Norfolk and Norwich Region formulae (1) and (2) gave estimated effective population of 484 583 $\pm$ 838, based on about 41 addresses for a total of some 60 urban and rural districts, and m cipal and county boroughs. The actual population of the adminis tive area for the Norwich group was about 546 550—an overestir of nearly 13 per cent. These figures refer to the year 1951 and ma now be out of date. At all events the method is clear. Work of kind is worth repeating and could easily be streamlined using punc card records. Using the minimum data of age, sex, specialty, hosp and administrative district, an extremely valuable investigation c be made. We could begin to study variations in demand in s detail, relating age and sex structure to geographical area, as we finding out whether the several specialties were supplied by the s or different effective populations.

Another subsidiary investigation which could be undertaken at same time would be to add to each record card the approxin distance of the patient's address from the nearest available outpat department. Since most non-acute cases are admitted via such dep ments, I believe that this would shed much light on the extent to wh the true medical requirement—as opposed to mere current reques is masked by the right clinic being relatively inaccessible. This wo be revealed by a marked negative correlation between the ac demand and the distance to the nearest appropriate clinic.

Estimating the Actual Demand. So much for the populatio risk. The second main aspect of the problem is to gauge the ac demand in each specialty of the group of hospitals under considerat

is involves a relatively straightforward analysis of hospital records. e must, however, be careful to supplement admission (or discharge) ures by changes in the length of the waiting list, and it is of course ential for the latter to be accurate and up to date. This has been cussed at some length elsewhere,[2] to which reference should be made details. It can be shown that for theoretical purposes the problem est thought of as a standard application of queueing theory. Briefly, owing the demand and average length of stay, we can calculate the ne that ought to be spent on the waiting list for different numbers of ds provided.

In general it is sufficient to find the so-called *critical* number of beds ich would supply exactly the amount of hospitalisation demanded, d to add to this figure one or two beds plus twice the appropriate ndard error for safety. For a random demand with constant average el this rule results in average waiting times of only a few days. If a particular specialty d is the average demand per day and s the erage length of stay in days, estimated from samples of n and n' servations respectively, then the critical number of beds required is ven by c, where

$$c = ds \qquad \ldots.(4)$$

d

$$\text{var}\,(c) = c^2\left(\frac{1}{n} + \frac{1}{n'}\right) \qquad \ldots.(5)$$

The critical number of beds required per head of population is then

$$R = c/M \qquad \ldots.(6)$$

ere

$$\text{var}\,(R) = R^2\left\{\frac{\text{var}\,(c)}{c^2} + \frac{\text{var}\,(M)}{M^2}\right\} \qquad \ldots.(7)$$

Prediction for a Different Population. If on the other hand we ve to budget for an entirely new area, say the construction of a tellite town, we are chiefly concerned to predict the probable demand the population expected, say N_0. Assuming that it is legitimate, at st as a first approximation, to employ estimates derived from elsere, we should reckon the critical number of beds to be c_0, given by

$$c_0 = RN_0 \qquad \ldots.(8)$$

ere

$$\text{var}\,(c_0) = N^2{}_0\,\text{var}\,(R) \qquad \ldots.(9)$$

e value of R and its variance would be given by (6) and (7) applied the data from earlier studies.

Even when advising on recommendations for new hospital building an existing population it will usually be convenient to make quick tial estimates based on rough figures of population before embarking a more exact investigation. Obviously there are many different uations that can arise, depending on how much reliable knowledge already available, and much further work still remains to be done.

L*

It is worth drawing attention to a purely mathematical problem is, however, of considerable practical importance. The number of required to satisfy a given demand is based on the theory of a n server queue with random arrival and negative exponential se time. What happens to the average waiting time if the rate of ar varies, though remaining random in character? We know, for exar that the demand for medical beds is higher in winter than in sum The general consequence is an annual fluctuation in the length o waiting list. But we do not yet know how to calculate the effect of a fluctuation in advance, and this might account for the size of n waiting lists, most of which seem very much longer than woul expected from the simplest theoretical models.

Outpatients. Similar considerations apply to outpatients as v Again we are faced with estimating the demand and finding the ef tive population at risk. An important difference this time is that ser takes place in *batches*, each batch being the number of patients see each clinic. The calculation of time spent on the waiting list for appointment now involves a special modification of queueing theory

So far we are thinking of *new* patients who represent a basic f tuating demand. *Old* patients do not present quite such a pres problem, as their demand, arising as it does from new patients have already had consultations, is more stable, predictable, and u control. It would, however, be of considerable interest to carry c detailed study of the whole pattern of first and subsequent app ments with a view to understanding more about the relation betv supply and demand.

Emergency Beds. An important aspect of the inpatient problem is admissions to hospital are really of two main kinds: non-acute from the waiting list and emergency cases directly they arise. Hosp normally deal with this by reserving a certain number of bed emergencies. The salient feature of the situation is that emergency can be reserved only at the expense of waiting-list patients. W therefore faced with balancing two opposing factors: the rat emergency beds occupied (i) to those actually demanded, and (i the number actually reserved. At the same time we must take acc of the fact that when the demand rises above the routine allowan *must* be satisfied by, for example, the setting up of extra beds in ce wards. A special statistical investigation of this has been made by D. J. Newell,[4] who has shown how a satisfactory rational solutio the problem can be achieved.

It is important to take into account not only seasonal fluctuation the demand for emergency beds but also the difference betw Sundays and other days. For average daily demands between five thirty beds Newell showed that if we have available two more b than the average demand on any day then 95 per cent of emerge cases will immediately find a vacant bed awaiting them. If we pro four extra beds this percentage goes up to 97·5.

ppointment Systems. It is appropriate at this stage to make some rence to the design of appointment systems for outpatients. As this already been discussed in great detail elsewhere[5, 6] we need do more here than indicate the main lines of the research.

One must naturally distinguish between waiting on a waiting list for appointment—or alternatively waiting until the day of a booked ointment arrives—and waiting in the waiting room for a consulta- that will take place some time during the day. The first kind of ting can be controlled according to the theory we have already ussed. For a given average demand, the rate at which consultations st be provided follows automatically when we have decided what rage time on the waiting list is to be tolerated. The second kind of ting, on the other hand, is quite independent and is related to the angement of individual appointment times in a particular clinic.

n general it can be said that the very considerable waiting ex- ienced by many patients in some clinics is largely due to the sultant heavily over-insuring against the possibility of his wasting a minutes. The consultant's time is obviously very valuable, but so hat of industrial workers, office employees and housewives. Some d of a balance must be struck, and according to our judgement of tive value so we can choose an appropriate appointment system. e scheme outlined below seems likely to be somewhere near optimum. As expected on common-sense grounds, excessive waiting is avoided staggering the patients' appointment times and by not having too ny present when the consultant starts seeing them. Indeed, many ics normally function along these lines, having good practical sons for doing so. It is, however, worth calculating what we should ect to happen under various conditions—partly to predict what ght happen if something different were tried and partly in the hope persuading certain hospitals to adopt more rational appointment ems.

To take a typical arrangement, suppose we have a clinic of twenty- patients. The total consultation time has a χ^2 distribution (with five d.f.) and a mean value of 5 minutes, so that the session will on average about $2\frac{1}{4}$ hours. Then a good plan is to give the patients ointments at 5-minute intervals and to have the consultant start rk as the second patient arrives. With this scheme it has been shown t the patients' average waiting time is about 9 minutes, and time sted by the consultant occasionally waiting for the next patient to ve averages approximately 6 minutes per clinic. In practice this e period' is likely to be largely eliminated by some patients arriving ly and by the consultant being able to fill in with a certain amount navoidable clerical work.

Actually, with this type of appointment system the patients' average ting time is shorter than the overall average at the beginning of the ic and longer at the end, the first and last patients having waiting es of 5 and 12 minutes respectively. This might be unfair to patients

with repeated visits if these always tended to occur near the end of clinic. It is possible to even things up by starting with 3 pa present and making the appointment interval about 8 per cent gr than the average consultation time. This, however, is a refine that may not be thought worth while in practice.

Factors influencing Internal Design

We shall now examine a number of factors that have a sp influence on internal design. Of course, those already discussed a major impact—the demand to be satisfied is immediately relate the number of beds that should be provided; or to the numbe clinics to be held, and hence to the size of outpatient departm Even appointment systems affect the latter, since the length of queue of patients, with their escorts, dictates to a great extent the of waiting space required.

However, supposing that the general scale of the facilities to be p vided has been decided, we next have to determine *how* this shall done. To a large extent this is a purely architectural matter. But worth asking whether any particular architectural solution is in fulfilling the required functions efficiently. We can also ask what be done to discover better solutions. This is therefore a typical fiel study for the methods of operational research.

There are special difficulties in trying to gauge the efficiency building, but it is often possible to study a number of factors walking distances, space used, light intensity, patterns of movem services required, etc. We can then try to calculate what var alternative designs suggested by architectural intuition are likel achieve. This gives a method of discriminating between the alternat Moreover, the application of such methods will itself have a modif influence on the architects' way of working.

For an extensive account of what can be done reference shoul made to the full Nuffield report already mentioned. We shall here at only a few illustrations, in which the mathematical and statis aspects are more immediately obvious.

Internal Traffic Studies. First let us take some of the problems invo in assessing the internal traffic of a hospital. At its simplest we thinking merely of the movement of human beings and various mat objects about the hospital.

The general principle to be followed is that areas with a l volume of mutual traffic should be sited near together, while those w less traffic may be more distant. This requirement may be modified questions of urgency or administrative convenience. There are main aspects of the problem to be distinguished. First, there is *total flow* of traffic between any two places, which we try to minim so as to cut down walking distances for both patients and staff, w consequent reduction in dirt and noise. Secondly, we try to minim the *traffic density* at any point at a given time in order to reduce l

gestion and ease the actual flow of traffic. Density is more difficult measure than total flow but can be measured approximately by mating the latter over short periods of time.

Traffic Surveys. When contemplating the construction of a new pital one of the main problems is to be able to make rough predic- ıs of the amount of traffic expected. There are two main ways of xling this. In the first place we can make surveys of existing pitals to discover what traffic actually occurs. This requires large mbers of observers stationed at strategic points to record the traffic. general it is best to do this so that individual journeys from one jor area to another can be distinguished. Since hospitals are all le to be rather individual, both in relation to local needs and to iation in organisation and administration, it is difficult to establish ad generalisations. However, analysis of data collected, combined h general experience, suggests that there are departments like out- ients, records, and X-ray that have high mutual traffic and should refore be closely sited. On the whole there appear to be no real ffic difficulties in hospitals apart from those involving outpatients, , and visitors. Detailed and costly traffic surveys are unlikely to be th while on a large scale.

Priori Predictions. This brings us to the second method of making dictions. Even if most hospitals in fact avoid serious difficulties, a experimental design might easily overlook some unforeseen com- ation. We are chiefly concerned to avoid bottlenecks and awkward rsecting lines of traffic. The suggestion is that proposed plans should carefully scrutinised with regard to probable traffic, especially that to outpatients, lifts, and visitors. To a large extent we can do this *iori.* Thus the volume and pattern of outpatient traffic can be ved from the estimates of demand already discussed, and the con- ion likely to arise from alternative clinic programmes can be ulated if the approximate sizes of the clinic sessions are known. traffic due to inpatients' visitors is known to be roughly two per ent, so the consequences of any particular visiting hours can be mated. Lift traffic is more difficult to deal with, and most of the mercial rule-of-thumb methods for calculating lift accommodation n floor area have no rational basis and are in any case unlikely to k for hospitals. This problem is certainly worthy of further study. most that can be said at present is that if ordinary lifts are large ugh and sufficiently numerous to deal with bed traffic, the con- ance of ordinary passengers is unlikely to give any trouble.

his means that though it may be difficult to calculate exactly in ance the best plans that could be devised, estimates of probable fic provide a useful critique of any actual proposal.

urther Problems. All this is relatively straightforward, but there are y other aspects of traffic in a wider sense that I believe might have important consequences for hospital design, if they were suffi- tly studied. We can broaden the concept of traffic to include not

only the movement of human beings and materials, but also ge
communications and storage of information. Pedestrian traffic w
certainly be reduced by better telephone facilities, call systems, and
adoption of more highly developed recording and filing appar
Such questions are especially important in any building where adm
strative, operational or technical efficiency are paramount. It s
likely that substantial progress would result from detailed cyberne
investigations, but these are beyond the scope of the present discuss

Ward Layout. While the previous discussion has been concerned w
traffic between the various main units and departments of a hosp
there is also the problem of movement and layout within each u
Take a ward unit, for example, that is the bed area plus the var
ancillary rooms. Then one of the aims of good design is to make
plan as compact as possible without overcrowding. For this a meas
of overcrowding is required. One method is to observe the patter
journeys between the main items, e.g. beds, duty-room, sluice-ro
etc., for a sample of nurses. This provides a *standard pattern of movem*
obtained by averaging over the different nurses. Thus for any act
ward unit or proposed design we can apply the standard. The lay
in question involves certain definite distances for each basic type
journey. Therefore, knowing the latter's standard frequency of oc
rence, we can calculate the total distance walked during, say, a si
spell of duty. This total distance is then a measure of compactness,
can be used to compare alternative layouts.

Space round the Bed. While on the subject of movement in the w
it is worth drawing attention to a very useful piece of work on the
space is used round a hospital bed. There are many factors invo
here, but we shall mention only one aspect: how much space is actu
required for standard nursing procedures to be carried out with
inconvenience? The method adopted was to film a number of
cedures and record the time that nurses used areas at various dista
from the bed. A frequency distribution based on the relative propor
of the time that these distances occurred then gave the main res
As usual in such cases one draws the line somewhere in the tail of
distribution, and might use a distance that was adequate say 95
cent of the time. Actually in bed-making 86 per cent of the time
spent within 2 feet of the bed and 98 per cent within $2\frac{1}{2}$ feet. For m
elaborate procedures up to 4 feet was required, but very rarely m
than this. For beds 3 feet in width this gives a minimum distanc
7 feet between the centres of adjacent beds. This figure is of co
based on movement alone: considerations of cross-infection m
demand something greater. Such an approach is simple, direct
effective. It could if necessary be extended to many other activit
and is clearly capable of more sophisticated applications than the
described here.

Single Rooms. Another important problem in ward design is to kn
how many beds should be accommodated in single rooms.[7,8] Th

several reasons for wanting to do this—some medical, some social. all events there is a demand for this type of accommodation which ies from day to day, and it is reasonable to assume a Poisson distribu-n. As with the somewhat analogous situation in the provision of ergency beds there are two kinds of efficiency between which some d of a balance has to be struck: (i) the *efficiency of provision*, i.e. the ent to which actual demand can be met; and (ii) the *efficiency of isation*, i.e. the extent to which single rooms are used for their proper rpose. The actual demand can be estimated from suitably designed veys of patients. A typical result obtained from applying the theory hat in an average 16-bed surgical unit two single rooms would sfy 84 per cent of the most urgent demand and the rooms would be d for this purpose 73 per cent of the time. If we go up to three gle rooms, the first percentage rises to 92 but the second falls to 59.

Early Ambulation. One of the most difficult things to allow for is the ct of changes or differences in medical technique or policy on ward ign. The importance of this is easily seen from the effect of the portion of ambulant patients in a ward. If most are bed-fast then need for day spaces and lavatories is reduced to a minimum. If the other hand a large number of patients are ambulant, then these ilities must be provided to a much greater extent, with consequent ercussions on the design. Goodall[9] studied 155 patients in surgical ts, and estimated how the proportions of patients in the main egories of bed-fast, partially ambulant, and fully ambulant would y from the actual regime (i) with a more traditional policy and (ii) h a definite bias in favour of early ambulation. The figures are wn below.

	Bed-fast (%)	Partially Ambulant (%)	Fully Ambulant (%)
Traditional	67	17	16
Actual	50	20	30
Early ambulation	16	35	49

There are certain obvious difficulties in interpretation, but such data quite useful as they stand. For a ward of given size one can calcu-approximately the numbers of patients likely to need the facilities ntioned, and the appropriate modifications can be incorporated in design. From a long-term point of view it might be well worth le making periodical surveys of medical opinion in order to gain ie idea of trends. After all, hospitals have to stand for a long time, a marked trend towards, say, earlier ambulation could easily make ard unit, adequate on present standards, awkward to run in a few rs' time.

Administrative Use of Current Statistical Returns

We have now seen how the right sort of statistical data—eith
the form of existing records or obtained from new surveys—will e
us to decide *how much* hospital accommodation is required. We
also seen how a good deal of the structure required can be investig
by the usual methods of operational research, using techniques
have a predominantly statistical basis. The final aspect to be disc
is keeping a check on the performance of existing hospital faciliti

In most separately distinguishable units of a hospital there mu
course be the ordinary routine records that enable the staff to do t
jobs properly. It is not with these that we are concerned here,
rather with those statistical returns that help to provide an ov
picture of how a hospital is broadly meeting the demand for med
advice and treatment. This is rather similar to the role played
quality control in industry.

Inpatient Departments. Consider first a given inpatient departm
say general medicine. A certain number of beds are allocated to
specialty and a definite proportion may be specially reserved
emergency admissions. We can now examine the kind of data alre
referred to as necessary for estimating demand: admissions or discha
and deaths, waiting-list figures and average lengths of stay, doing
month by month, or quarter by quarter. Although these are
sufficient for relating supply and demand, it is advisable to have a
more to assess the actual efficiency of the supply. Bed-occupan
a favourite index but is liable to be misleading unless carefully rel
to average length of stay. A more immediately usable statistic is
average turnover interval, i.e. the time that a bed is vacant bet
successive patients. When there is a waiting list the turnover inte
is a direct measure of one aspect of efficiency.

If we inspect such tables and compare the results for various hosp
at different times, we can with a little practice get quite a good id
how things are going. The picture is not of course complete. But
marked divergencies in turnover intervals, changes in waiting list,
can easily be picked out for discussion. Sometimes local knowl
will immediately supply the explanation; at other times specia
vestigation will be needed. This sort of thing is regularly done by
Oxford Region, for example, and proves extremely useful. One o
chief advantages of a current appraisal is that any unsatisfactory tr
are likely to be picked out sooner than would otherwise be the c
Moreover such data are essential if we are to use queueing theor
calculate the probable consequences of emergency measures such
putting up temporary extra beds to work off an unduly long wai
list. It goes without saying, however, that every effort should be m
to have accurate figures—waiting lists, in particular, are often diffi
to keep properly up to date.

Outpatient Departments. Similar methods can clearly be applie
outpatient departments. A regular check on current demand

ting times shows very quickly whether anything unexpected is ng place. An additional possibility here is to keep account of the rent situation with regard to the functioning of appointment ems. In particular, we can record patients' waiting times, not getting their punctuality and rate of defaulting!

General Value of Current Checks. The general value of such current cks should be obvious. Care is needed in the interpretation of the e figures, which though a very important guide are not the whole y. The question of the medical quality of the services supplied is which must constantly be borne in mind and is a difficult matter ake into account. No doubt the judgement of the medical staff is newhat surer when the basic facts are more certain. But what sibility there is of improving the purely medical assessment of lts is a subject which requires further attention, e.g. how does one ide whether patients in one surgical unit with a short length of stay being treated better or worse than those in another unit with ger stay?

Hospital records must of course be sufficiently detailed to meet ical requirements. They should probably be streamlined to an nt that just allows them to produce without difficulty the running cks mentioned above. A greater degree of complexity is unlikely to economical, and as already emphasised elsewhere,[10] I believe that development of highly complicated and extensive record systems nded for retrospective research is an entirely forlorn hope.

nowledgement

t is a pleasure to acknowledge my special debt to the Research Team he Investigation into the Functions and Design of Hospitals, spon- d by the Nuffield Provincial Hospitals Trust and the University of tol, for first encouraging my interest in the problems dealt with in paper and for many stimulating discussions.

REFERENCES

UFFIELD PROVINCIAL HOSPITALS TRUST (1955). *Studies in the Functions and Design of Hospitals.* Oxford.

ILEY, N. T. J. (1954). 'Queueing for medical care', *Applied Statistics*, **3**, 137.

m. (1954). 'On queueing processes with bulk service', *J. R. Statist. Soc.*, B, **16**, 80.

WELL, D. J. (1954). 'Provision of emergency beds in hospitals', *Brit. J. Prev. Soc. Med.*, **8**, 77.

ELCH, J. D. and BAILEY, N. T. J. (1952). 'Appointment systems in hospital out-patient departments', *Lancet*, **1**, 1105.

ILEY, N. T. J. (1952). 'A study of queues and appointment systems in hospital outpatient departments, with special reference to waiting times', *J. R. Statist. Soc.*, B, **14**, 185.

ODALL, J. W. D. (1951). 'Single rooms in hospital: estimate of the medical need', *Lancet*, **1**, 1063.

ILEY, N. T. J. (1951). 'On assessing the efficiency of single-room provision in hospital wards', *J. Hyg., Camb.*, **49**, 452.

ODALL, J. W. D. (1951). 'Early ambulation: a survey of hospital practice', *Lancet*, **1**, 43.

ILEY, N. T. J. (1952). 'The scope of medical statistics', *Applied Statistics*, **1**, 149.

MONTE CARLO METHODS AND INDUSTRIAL PROBLEMS

W. NEIL JESSOP
Operational Research Department, Courtaulds Limited

Although the Monte Carlo method is not new, it has re achieved a good deal of prominence. In this article Mr J shows, by means of two examples, how the technique m applied to solve industrial problems, the mathematical so of which is difficult or impossible.

In a recent paper[2] Hammersley and Morton concisely define the M Carlo method as one 'which solves problems by subjecting random n bers to numerical processes'. Although it has only recently acquire present nickname, it has been in use for many years for the solutio mathematical problems by their statistical analogues. Nuclear blems now seem to have stimulated a new interest in the method to have given impetus to its development as an efficient techn Since the method is essentially solution by a sampling experime might justifiably be said that mathematical statisticians have using it for a long time, for example in the determination of sam distributions and the power of tests. In recent years operat research workers too have found uses for the method, particula queueing problems, where the situations are often so complex t mathematical solution is virtually impossible. It is probable, how that the method is still not assessed at its proper value by ap statisticians, and this article has therefore been prepared to stim interest in it. In the following sections the essentials of the metho be outlined and illustrated by examples of a relatively simple from an industrial background.

Outline of Method

The Monte Carlo method is one where the statistician himself data-producing agent, and in which inferences are made from sa drawn from a 'statistical model' of the situation under investiga By 'statistical model' is meant the set of assumptions concernin behaviour and constraints of the system, and the statistical vari concerned. Thus, in a queueing problem the model might concer input and servicing distributions, queue discipline, and assump concerning 'feed-back', e.g. when the queue length is greater th certain number, additional servicing facilities become available. a situation can be artificially realised on paper in terms of data fact, when the model is fixed, the next step is to draw up the data with particular regard to ease of working and comprehensibility

in data one has recourse to a table of random numbers, and the edure is as follows:

ıppose that the model entails numerical operations with sample es from three distributions of which the first is defined by a cumula- distribution function of $\phi(x)$, $0 \leqslant \phi(x) \leqslant 1$; a number, say 78, is vn from the table, then the variate x is chosen such that $\phi(x) = 0{\cdot}78$, similarly for the remaining two distributions. As a general rule it visable to 'embed' the random variates in the data sheet before the ımetic and other operations are carried out, which implies, first of ampling from the distributions concerned to supply a good run of erial for calculation. At various stages in the work the quantity in tion may be estimated with its standard error, so that a decision be made at any stage whether or not to stop. In the application e method to industrial problems labour can be carried to excess chieve fictitious accuracy, since the underlying assumptions may ustify it. This point is referred to later, but this would seem to be ssential difference between operational research and exact science ications, and deserves stressing here.

nple 1: *Plant Failure Problem*

factory intends to build a plant consisting of two sets of three eaters feeding into a common kiln, each set of preheaters being ɔsed in a furnace. The arrangement is illustrated in Fig. 1.

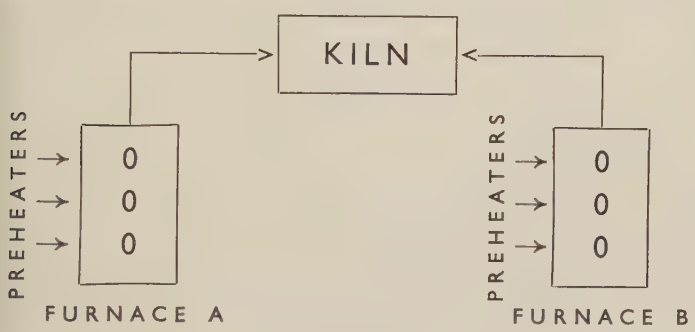


Fig. 1. Layout of preheater plant.

e failure of any one preheater within a furnace will cause the lrawal of the whole furnace from production for a period of time to twenty days, during which time the remaining furnace will be ɔaded if production is to be maintained. During the period of r all three preheaters are renewed. More serious dislocation will however, if one furnace should fail during the repair time of the Provision must be made for such an event, but it is undesirable his should be of too frequent occurrence. What are the chances e or more 'double failures' within six months or a year? It is that the length of a repair period affects the answer to this on, so that it will be necessary to examine a range of repair ls from the shortest feasible to the probable maximum.

The data consist of lives of individual preheaters referring different technical set-up, and a basic assumption is that the working conditions are not sufficiently different to affect the lif tribution. This distribution is shown in Fig. 2, and is based on a siderable volume of informa It does not conform to any known or analytically manag distribution. The problem is very well suited to the Monte method.

FIG. 2. Frequency distribution of lives of preheaters.

On the assumption of the val of this life distribution, data r ing to the proposed plant ca generated in a simple manner furnace will fail if any one of three preheaters fails, so tha life will be that of the shortest-l member of the set. In terms o Monte Carlo model, therefore, the furnace life is given by the variate of a set of three drawn at random from the distributio Fig. 2. The manner in which this is done is as follows. Succe digits in a random number table are 98, 12, 70; the least ran number will correspond to the least variate, so that 12 is chose the cumulative form of the distribution, the life, such that 12 per of the population are less than it, is 86 days, which is therefor furnace life. A series of lives for a furnace are obtained from a tal random numbers and the cumulative form of the distribution. lives must be converted into instants of failure, and this is dor adding the fixed repair period on to each life (except the first) finding successive cumulative sums. The method is illustrated b beginning of the data sheet for a repair period of sixteen days (Tab

Furnace A data are first obtained, the instants of failure being in the second column. A series of lives for furnace B is obtained b

TABLE I

Method of estimating double failures from Monte Carlo model; repair period 16 da

Furnace A		Furnace B		Double Failure
Life (days)	Cumulative	Life (days)	Cumulative	
86		73		*
132	234	227	316	
172	422	175	507	
132	570	184		
60	646			
58	720		707	*

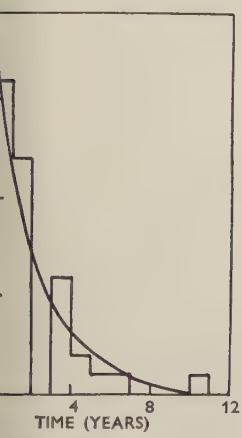
e process and the cumulatives are formed with an eye to the figures ady entered for furnace A, so that if a number differs by less than rom a furnace A cumulative it is aligned accordingly, and the ble failure noted by an asterisk in the last column. Repair periods 12, 16, and 20 days were taken. Sufficient random numbers were to give 400 failures for each repair period.

TABLE II

Calculated probability of r double failures in 180 *days*

r	Repair Time (days)			
	8	12	16	20
0	0·845	0·811	0·784	0·729
1	0·142	0·170	0·191	0·230
2 or more	0·013	0·019	0·025	0·041

ne data for each repair period were studied with regard to the vals of time between double failures, i.e. asterisks on the data t. It was found that the distribution of this time interval conformed closely in all cases with the exponential, a distribution wholly rmined by the mean interval between double failures. The latter meter depends on the repair period, as would be expected; for nple, it was 570 days for a 20-day repair period and 1 070 days n 8-day repair period. The distribution for a 16-day repair period is shown in Fig. 3. Since the exponential distribution is such a good fit, the remainder of the problem can be solved by standard probability theory. Thus, the mean number of double failures in a period T is $T/\bar{t}$, where $\bar{t}$ is the mean time interval between double failures. Probabilities of 0, 1, 2, . . . double failures in any period can be calculated using the Poisson distribution. In Table II probabilities are given for repair periods of 8, 12, 16, and 20 days and $T = 180$ days $\simeq$ 6 months. For all repair periods the risk of 1 double failure is appreciable, but the risk of 2 or more is small.



Frequency distribution of nterval between double s of preheaters (16-day period), with fitted exponential curve.

This example is typical of a class of problems occurring in industry concerned with plant replacement and maintenance. Some are soluble by analytical processes (see r[1] on renewal theory, for example). However, when actual conns like the deployment of maintenance gangs and real plant re behaviour are considered the applicability of the Monte Carlo od becomes clear.

Example 2: *Machine Servicing Problem*

A new chemical process is to be carried out in vats. Work asso with the process consists of emptying each vat and charging emptying can be regarded as ir taneous, but the charging is a plex of manual operations v distribution of a very skew known from similar process shown in Fig. 4. It is proposed one man in charge of ten arranged in a line. When a sig visible, he is to move along th of vats emptying and charging manner illustrated by Table I

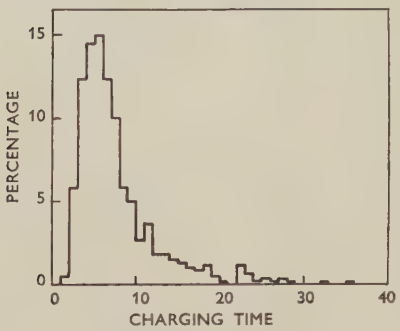


FIG. 4. Frequency distribution of vat-charging time.

The instant the signal show 1 is emptied and at the end units of time charging is complete; vat 2 is emptied the moment is charged, and if charging then takes t_2 units it is complete at t_1 charging and emptying proceed down the line in this manner completion of charging on one vat coinciding with the emptying next.

TABLE III

Timetable for emptying and charging vats

	Vat 1	Vat 2	Vat 3	Vat 4
Empty	0	t_1	$t_1 + t_2$	$t_1 + t_2 + t_3$
Charge complete	t_1	$t_1 + t_2$	$t_1 + t_2 + t_3$	$t_1 + t_2 + t_3 + t_4$

The mean reaction time depends on the interval between succ signals. What must the time interval between signals be to g required mean reaction time T? What variation about this reaction time may be expected?

The first question can be answered directly. Consider the ser sequence that is the successor to the one illustrated in Table III. interval between signals is τ, the corresponding times for emptyi τ, $\tau + t_1'$, $\tau + t_1' + t_2'$, and similarly for the charging times. the reaction times for vats 2 and 3 are respectively $\tau + t_1' - (t_1$ and $\tau + t_1' + t_2' - (t_1 + t_2 + t_3)$. Their mean values are ide and equal to $\tau - \bar{t}$. Thus for the mean reaction time to be T the interval (τ) between signals should be set such that $\tau = T + \bar{t}$.

The expressions for the reaction times also indicate that for vat 10 there will be distributions of reaction time with increasing var although with the same mean. What is required to answer the se question is knowledge of the distribution which is obtained by s posing all vat distributions. It is not easy to see how this can be de

ıematically from knowledge only of the empirical distribution of 4.

ıe layout of the Monte Carlo calculations is illustrated in Table IV.

TABLE IV

Examples of servicing sequences generated from Monte Carlo model

No.	S14			S15			S16		
	E	C	RT	E	C	RT	E	C	RT
1	0	13	−3	0	7	−13	0	5	−7
2	13	17	−2	7	11	−10	5	30	−6
3	17	35	−2	11	16	−24	30	33	14
4	35	38	13	16	20	−22	33	42	13
5	38	42	13	20	24	−22	42	47	18
6	42	47	14	24	37	−23	47	58	10
7	47	51	11	37	52	−14	58	67	6
8	51	54	12	52	57	−2	67	72	10
9	54	57	10	57	70	0	72	75	2
0	57	76	11	70	80	−6	75	81	−5

4, S15 and S16 are successive servicing sequences in the numerical el; C, E, and RT are abbreviations for Charge Complete, Empty, Reaction Time. A sheet of random sets of ten charging times is obtained by using random numbers in conjunction with the ging distribution in the manner described previously. Those for are 7, 4, 5, 4, 4, 15, etc., and column C shows the cumulative sum ese, which, for each vat, gives the instant that charging was com- for that vat and the instant the vat next in the line was emptied. RT column is obtained by taking the difference of any emptying in (say) S15 with the corresponding 'charging complete' time in All the RT values require the addition of the time between ls (τ). In the whole experiment 590 reaction times were generated is way. The resulting cumulative distribution is shown by the nuous curve in Fig. 5. In this case the 95 per cent limits for the ion time are from $\tau - 47$ to $\tau + 18$, a range of 65 units.

ppose now that it is considered that this range of variation is too . One way of reducing the variation is to employ two men on the cing. Therefore, suppose that when a signal shows, one man nences work on vat 1 and another commences work on vat 6, eeding down the line to vats 5 and 10 respectively. A distribution for this method of working is obtained, using the same data s, simply by considering the reaction times for vats 1 to 5 only, s shown in Fig. 5 as the broken curve. The range of variation is

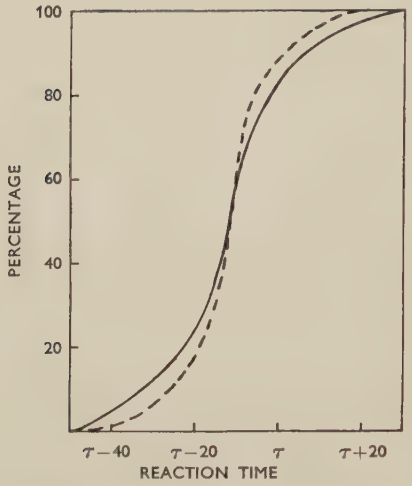


FIG. 5. Cumulative distribution of reaction time: continuous curve, one man; broken curve, two men.

now $\tau - 37$ to $\tau + 11$, 48 un substantial reduction.

The method may be used t timate the improvement that w result from better techniqu partial mechanisation in hand It may also give a guide to the size of labour force and its de ment in a factory departme which there is a variety of wo a similar type to that describe the example.

Discussion

Both of the examples are cerned with distributions. cases very often involve a l work, since to obtain a really knowledge of a distribution a sample size amounting to several thou must be generated. In the examples chosen this was not neces since in the first case the time intervals fitted a known distribution sufficient precision for inferential purposes, and in the second c was recognised that the charging-time distribution might not apply sufficient exactness, under the different conditions of plant layou visaged, to justify high precision in the Monte Carlo work. The case is analogous to the commonplace rule that the number of de places in arithmetic work must be justified by the underlying basis calculation. It is something which needs to be remembered here, the mathematical model is a set of assumptions concerning a situation which may in fact be very complex. This model may, ever, sometimes be tested directly. If recommendations suggest the calculations are put into effect, results can be compared with predicted. Weaknesses in the assumptions may thus be broug light so that another model may be set up, which in turn may better agreement, and serve as a sounder foundation for other re mendations. The Monte Carlo approach is thus seen as a parti expression of the scientific method.

The problem of the amount of Monte Carlo sampling required the solution of a problem and the whole question of the most effi design (in terms of accuracy and time) are receiving much attentio present, and for a discussion of the problem the reader is referre the paper by Kahn and Marshall.[3] One method given in this p is of particular interest and deserves special mention. The situa might arise where the problem is a comparison of two policie methods of working. There are thus two variables x and y and int is centred on $z = x - y$, its mean and standard error. The techniq analogous to the paired t-test and is called the 'correlated samp

hod. Since var (z) = var (x) + var (y) − 2 cov (x, y), if cov (x, y) be made large and positive var (z) will be considerably smaller n if x and y were uncorrelated. This correlation may be obtained using as many of the same random numbers in paired repetitions ne alternative policies as possible. If the policies differ in nature it not be possible to do so to a sufficient extent, but if they are lar a significant reduction in sample size may be obtained. The dard error may be computed at various stages in the work and the pling discontinued when it is thought that the difference is known sufficient accuracy.

has been stressed elsewhere[1] how important it is that thought uld be given to the form in which the data are collected, the layout he worksheet, and the reduction of work to the simplest and most matic processes, with, if possible, provision for routine checking at erent stages. The examples cited here are so simple that they do make this point very strongly. However, industrial problems can difficult to the extent that pen-and-paper methods have to be ndoned and machines used instead. The discipline of organising to some extent carrying out the calculation can be of great service larifying the problem, and for this reason it is recommended that statistician in charge should not delegate work in its initial stages. ne problem is capable of an analytical solution this is more likely ecome apparent if a Monte Carlo approach is attempted than rwise. Statisticians should therefore not regard their use of this nod with shame, as a retreat after defeat, but rather as a healthy erimental approach, in the best scientific tradition.

owledgements

am indebted to Miss R. A. Hamer for help in preparing this paper, Mr F. I. Musk for the material of Example 1, to Mr A. J. H. ley and Mr A. W. Swan for constructive criticism, and to the ctors of Courtaulds Ltd. for permission to publish.

REFERENCES

LER, W. (1950). *An Introduction to Probability Theory and its Applications, Vol. I.* Wiley, New York.

MMERSLEY, J. M. and MORTON, K. W. (1954). 'Poor man's Monte Carlo', *J. R. Statist. Soc.*, B, **16**, 23.

HN, H. and MARSHALL, A. W. (1953). 'Methods of reducing sample size in Monte Carlo computations', *J. Operat. Res. Soc. Amer.*, **1**, 263.

E, R. R., BROWN, F. B. and BLANCHARD, R. O. (1955). 'An analysis of a railroad classification yard', *J. Operat. Res. Soc. Amer.*, **3**, 262.

, H. G. and LEE, A. M. (1955). 'Monte Carlo methods in heavy industry', *Operat. Res. Quart.*, **6**, 108.

THE ALLOCATION OF SCARCE MATERIALS BETWEEN PRODUCTS

NOEL WILLIAMS, FIA, ACWA

Mr Williams outlines the principles of Linear Program and shows, by a simplified example worked out in detail, it can be applied to problems of material allocation.

In a manufacturing industry in which the final products are mad combining and processing a number of raw materials, a company m be able to obtain all that it requires of certain materials but c limited quantities of others. These other materials may not be exp sive if they are by-products from other industries, but the quanti available will depend on the output of the principal products of th industries. A company which can use these by-products as raw ma rials for its own products must decide how much to buy of each decision should be based on the prices, qualities, and availabilitie all the raw materials which it could use, the extent to which these products could be used to replace the main raw materials, the capa of its plant or outside firms to perform any treatment on these products necessary to make them suitable, and the cost of such tr ment. If these by-products are seasonal, then it is also necessar decide when these materials should be used; and this decision sh take account of the expected future prices of all materials, the co storing the by-products, and any limitations on storage capacity. mathematical technique of Linear Programming provides a metho working out the optimal buying policy, taking these factors account. The calculations would show how much of each mat should be purchased, how much should be used in each product, if the time factor is taken into account, when it should be used.

In the previous paragraph the 'optimal' buying policy was m tioned. This points to one essential feature—there must be alternat and there must be some numerical assessment of each alternative so there is some way of defining which is the best. These alternatives not straightforward *either-or*'s, but are combinations of indivi elements (known as 'activities') which can be combined in var amounts and in various permutations to give alternative soluti The number of possible combinations is so great that it is not practica to work out all possible solutions and evaluate each, and thus pick best.

The next feature is that the relationships between the elements their measurements must be linear. Another important featur divisibility of the elements. For example, consider a ten-ton l

ying sand; the load is divisible, up to a maximum of ten tons, as can carry any quantity up to that amount, but the lorry is not sible—we must run the whole lorry whatever the load. This some-es leads to practical difficulties.

One of the first steps in considering the application of linear pro-mming is to examine the problem in detail and enumerate the rations: these operations, which are generally known as 'activities', st be rigidly defined. An 'activity' is a very general term used to cribe some action such as, for example:

a) consuming a specified raw material;
b) travelling by a specified route from one named place to another;
c) using a particular manufacturing process;
d) buying a specified product;
e) holding a specified product in stock for a specified period; and so on.

h activity is given a unit level; the above activities might have the owing unit levels:

a) one ton of the raw material;
b) one journey;
c) one ton of input or one hour running;
d) one pound of product;
e) one ton for one day.

h activity has a unit value which is the value of using the activity nit level, and may be a cost, a profit, a distance, an hour, etc.

The fundamental basis of the calculations is this: having expressed the operations as a set of activities, we put these activities and the uirements into a system of equations; we find a solution which ts all requirements (this is a feasible solution); and then, by a es of logical steps, we improve upon this feasible solution until we ain the optimal solution.

To take a simplified example of the more complicated problems met practice, suppose a company wishes to manufacture 100 units of n of two products, R and S, where each unit is made from 1 ton of material; product R is made of material A costing £12 per ton and duct S of material B costing £8 per ton; each product could con-up to 5 per cent of a third material (a by-product) D costing £2 ton. If all three materials were available at these prices and in cient quantity, then evidently the company would make 100 units from 95 tons of A plus 5 tons of D and 100 units of S from 95 tons plus 5 tons of D. If less than 10 tons of D were available, then it ld be used up to 5 tons in R and any balance in S—that is to say, ould be used first in that product where it replaces the most expen-material.

We can put the above problem into mathematical form by building a system of linear simultaneous equations; the first step is to specify

a number of activities; in this case we will define an activity as a specified material in a specified product, i.e. IJ is the activity of material I in product J. There are four activities:

Activity AR: Using material A in product R
Activity BS: Using material B in product S
Activity DR: Using material D in product R
Activity DS: Using material D in product S

For each activity we specify a unit level; in this case a convenient level for each activity would be 1 ton. The symbol x_{IJ} denotes level at which activity IJ is actually used, i.e. x_{IJ} denotes the nun of tons of ingredient I used in product J, and in our example denotes the number of tons of ingredient A used in product R.

We can now build up the equations and inequalities which, w solved, will give us the optimum allocation.

E1: We must use 100 tons of materials in product R, hence

$$x_{AR} + x_{DR} = 100$$

E2: We must use 100 tons of materials in product S, hence

$$x_{BS} + x_{DS} = 100$$

IE3: We cannot use more than 5 tons of D in R, hence

$$x_{DR} \leqslant 5$$

IE4: We cannot use more than 5 tons of D in S, hence

$$x_{DS} \leqslant 5$$

IE5: The maximum quantity of D which is available in the period tons, hence

$$x_{DR} + x_{DS} \leqslant \mathcal{N}$$

The next step is to convert the inequalities into equations by in ducing additional unknowns (slack vectors), here denoted by q's follows:

E3: $x_{DR} + q_1 = 5$

E4: $x_{DS} + q_2 = 5$

E5: $x_{DR} + x_{DS} + q_3 = \mathcal{N}$

These q's represent the difference between the level at which activity or combination of activities is actually used, and the upper lower) limit of that activity or combination of activities. The ab system gives five equations (E1 to E5) in seven variables (four x's three q's), subject to the restriction that no x and no q can be nega (a negative q means that we have exceeded some limit). As it sta this system of equations is indeterminate, having an infinity of soluti so we need some criterion by which to determine which solution is one we require. A satisfactory criterion in this case would be that total material cost should be as small as possible.

c_{IJ} denotes the cost of using activity IJ at unit level the total cost ır activities will be $\Sigma(c_{IJ} \cdot x_{IJ}) = \mathcal{Z}$. The quantity $\mathcal{Z}$ (known as the ctional') is to be minimised. In our example $\mathcal{Z} = 12x_{AR} + 8x_{BS} +$ $+ 2x_{DS}$ (the q's are taken as having a nil cost), where the coeffi- ts are the costs per ton of the materials.

t this stage it is worth distinguishing between limits which may be tive and those which cannot. If $\mathcal{N}$ (the limit of E5) is less than 10 ay be effective, as then we will not be able to use the full 5 tons allowed in both R and S. If, however, $\mathcal{N}$ is greater than 10 this cannot be effective, because the other limits restrict the maximum e of D to 10 tons altogether; hence in such a case E5 could be ted from the matrix. Limits which cannot be effective can usually ound while the problem is being investigated; but if they are not ed, and are included in the matrix, it does no harm (apart from all increase in the calculating time) and the same optimum solution be obtained. For example, if $\mathcal{N} = 14$, and we obtain $x_{DR} = 5$, = 5, then $q_3 = 4$, being the difference between the quantities used the limit of 14 tons available.

practice, problems of allocating scarce materials can be more plicated than the simple example discussed above in at least four :

The number of products may be larger; their specifications may nore complicated and may include attributes included in the rials in different amounts.

) The number and variety of types of ingredients may be larger, some ingredients may themselves be combinations of ingredients veral types.

i) The materials may be capable of up-grading by treatment re manufacture, there may be limits on the quantities that can be ed, several methods of treatment may be available, and the us treatments may cause different process losses for different raw rials.

) The possibilities of incorporating several time-periods into the ix may have to be considered when the supplies of materials and/or production programme are irregular, and this will introduce ge facilities and limitations.

hatever the nature of the limitations and restrictions, we build up tem of linear simultaneous equations and/or inequalities (on the of E1 to E5 above), and obtain the appropriate functional, $\mathcal{Z}$. mathematic model will then have the general form

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ . \ . \ . \ + a_{1n}x_n \leqslant \text{or} = \text{or} \geqslant b_1$$
$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \ . \ . \ . \ + a_{2n}x_n \leqslant \text{or} = \text{or} \geqslant b_2$$
$$\vdots$$
$$_1x_1 + a_{m2}x_2 + a_{m3}x_3 + \ . \ . \ . \ + a_{mn}x_n \leqslant \text{or} = \text{or} \geqslant b_m$$

subject to every x being positive or zero but never having a ne value.

If any of the above expressions are inequalities, then we c them to equations by introducing additional variables (slack ve one additional variable for each inequality, and the general becomes

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ . \ . \ . \ + a_{1n}x_n \quad (\pm q_1) &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \ . \ . \ . \ + a_{2n}x_n \quad (\pm q_2) &= b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \ . \ . \ . \ + a_{mn}x_n \quad (\pm q_m) &= b_m \end{aligned}$$

with the associated functional

$$c_1x_1 + c_2x_2 + c_3x_3 + \ . \ . \ . \ + c_nx_n = Z$$

In the course of our calculations we work out the economic val using each activity at unit level; this value, for activity j, is den by Z_j. We compare the economic value (Z_j) with the actual co using that activity at unit level (c_j); the difference $(Z - c)_j$ s whether, at its price, an activity could profitably be used.

In its general form the system of equations contains more vari than equations. If the number of variables equalled the numb equations there would (normally) be one solution; as we have variables than equations we have to choose m of them (m bein number of equations), and then, by putting all the other var equal to zero, solve this system to obtain its unique solution. T variables chosen are called the 'basis' variables; the others are the 'non-basis' variables. We use the values of $(Z - c)$ to d whether to change the set of basis variables by substituting one non-basis variables for one of the basis variables.

If we have some system of simultaneous equations the solution normally obvious, but it can be made so by putting it into a sp form.

Let

$$ax + by + cz = h \quad \ldots$$

$$dx + ey + fz = k \quad \ldots$$

Divide (1) by a [the coefficient of x in (1)]:

$$x + \frac{b}{a}y + \frac{c}{a}z = \frac{h}{a} \quad \ldots$$

Multiply (3) by d [the coefficient of x in (2)]:

$$dx + \frac{bd}{a}y + \frac{cd}{a}z = \frac{dh}{a} \quad \ldots$$

uct (4) from (2):

$$\left(e - \frac{bd}{a}\right)y + \left(f - \frac{cd}{a}\right)z = \left(k - \frac{dh}{a}\right)$$

may be written $ry + sz = p$ ....(5)

e have transformed our original pair of equations (1) and (2) into w pair: (3) which replaces (1), and (5) which replaces (2). The icient of x is 1 in (3) and 0 in (5).

de (5) by r [the coefficient of y in (5)]:

$$y + \frac{s}{r}z = \frac{p}{r} \quad \ldots.(6)$$

iply (6) by b/a [the coefficient of y in (3)]:

$$\frac{b}{a}y + \frac{bs}{ar}z = \frac{bp}{ar} \quad \ldots.(7)$$

uct (7) from (3):

$$x + \left(\frac{c}{a} - \frac{bs}{ar}\right)z = \left(\frac{h}{a} - \frac{bp}{ar}\right)$$

may be written $x + tz = u$ ....(8)

ave now transformed equations (1) and (2) into

$$x + tz = u \quad \ldots.(8)$$

$$y + \frac{s}{r}z = \frac{p}{r} \quad \ldots.(6)$$

utting $z = 0$ we obtain at once

$$x = u$$

$$y = \frac{p}{r}$$

n gives a feasible solution of equations (1) and (2) with x and y as variables.

mmarising the equations in our original example, and putting the bles at the head of each column, so that we write only the cients, we obtain:

Matrix I

	x_{AR}	x_{BS}	x_{DR}	x_{DS}	q_1	q_2	q_3	
E1:	1	.	1	.	.	.	.	= 100
E2:	.	1	.	1	.	.	.	= 100
E3:	.	.	1	.	1	.	.	= 5
E4:	.	.	.	1	.	1	.	= 5
E5:	.	.	1	1	.	.	1	= 8

(assuming that $\mathcal{N} = 8$)

The five equations which make up the above matrix are alre the form of equations (8) and (6), that is to say, we can choose f basis the five variables x_{AR}, x_{BS}, q_1, q_2, and q_3, and by p $x_{DR} = x_{DS} = 0$ can easily obtain as a solution

$$\begin{aligned} x_{AR} &= 100 \\ x_{BS} &= 100 \\ q_1 &= 5 \\ q_2 &= 5 \\ q_3 &= 8 \end{aligned}$$

and $\mathcal{Z} = 12 \times 100 + 8 \times 100 + 0 \times 5 + 0 \times 5 + 0 \times 8 = 2$
The latter is the cost, in £'s, of this solution.

Write the basis variables in a column to the left of the matrix, head the final column P:

Matrix II—Feasible Solution

		x_{AR}	x_{BS}	x_{DR}	x_{DS}	q_1	q_2	q_3	
E1:	x_{AR}	1	.	1	.	.	.	.	=
E2:	x_{BS}	.	1	.	1	.	.	.	=
E3:	q_1	.	.	1	.	1	.	.	=
E4:	q_2	.	.	.	1	.	1	.	=
E5:	q_3	.	.	1	1	.	.	1	=

This solution is an answer to our original problem in that it r all the requirements, and is known as a 'feasible solution'; we now to decide if it is the best solution. To decide this we have to calc the economic values of DR and DS, the unused activities.

As D could replace A, it would be worth paying up to £12 pe for D; as D could replace B it would be worth paying up to £8 pe for D. The economic value of an activity is not generally so evi so we need some means of evaluating it, and we can use the matri doing so.

We calculate $\mathcal{Z}_{DR}$ (the economic value of activity DR at unit l in a similar manner to $\mathcal{Z}$. Using the column headed x_{DR} we have

$$\mathcal{Z}_{DR} = c_{AR}.1 + c_{q1}.1 + c_{q3}.1 = 12 + 0 + 0 = 12$$
(all costs are in £'s)

Similarly $\mathcal{Z}_{DS} = c_{BS}.1 + c_{q2}.1 + c_{q3}.1 = 8 + 0 + 0 = 8$

Because we could *replace* A or B by a similar quantity of D, the sa per ton replaced is the *difference* between the cost of what goes ($\mathcal{Z}_{DR}$ or $\mathcal{Z}_{DS}$) and what comes in to replace it (c_{DR} or c_{DS}). $c_{DR} = c_{DS} = 2$, hence $(\mathcal{Z} - c)_{DR} = 12 - 2 = 10$ and $(\mathcal{Z} - c)_{D}$ $8 - 2 = 6$. The largest positive value of $(\mathcal{Z} - c)$ is 10, so we will DR into the basis, and we now have to decide which of the five varia at present in the basis will come out. We make this decision by 'division test': divide every value in the P column by the correspon value in the x_{DR} column. We obtain 100/1 = 100 (row E1); 5/1

E3); 8/1 = 8 (row E5). (We do not make the test for any value
ne x_{DR} column which is zero or negative.) We select the smallest
e, that is row E3, against which the basis vector is q_1; this is the
able to go out of the basis. The coefficient of x_{DR} in equation E3,
ely 1, is the 'pivot', and the other values which appear in the
mn containing the pivot are the 'factors'.

s we reduce our total cost by £10 for every ton of D that replaces
n of A, we naturally wish to replace as much as possible. The
ion test is a method of testing the restrictions of the problem to
out which one will set the limit to the replacement. By taking the
lest ratio as our criterion we ensure that we do not violate any of
estrictions; for example, if we had chosen the ratio 8 and had tried
ut 8 tons of D into R to replace 8 tons of A we should have exceeded
tons the restriction that 100 units of R could contain not more
5 tons of D.

'e divide the equation containing the pivot by the value of the
t, i.e. divide E3 by 1; we then multiply the values so obtained by
factors (1 in E1 and 1 in E5); we deduct the values so obtained
the corresponding values in the rows containing the factors.
de E3 by the pivot, 1; we obtain

$$E3': x_{DR} + q_1 = 5$$

multiplied by the factor in E1, i.e. by 1, gives

$$E3'\ (1): x_{DR} + q_1 = 5$$

larly $$E3'\ (5): x_{DR} + q_1 = 5$$

factors in E2 and E4 are zero).

uct E3′ (1) from E1:

$$E1': x_{AR} - q_1 = 95$$

uct E3′ (5) from E5:

$$E5': x_{DS} - q_1 + q_3 = 3$$

marising the equations in Matrix III, remembering that x_{DR} has
aced q_1 in the basis:

Matrix III

		x_{AR}	x_{BS}	x_{DR}	x_{DS}	q_1	q_2	q_3		P
E1′:	x_{AR}	1	.	.	.	−1	.	.	=	95
E2:	x_{BS}	.	1	.	1	.	.	.	=	100
E3′:	x_{DR}	.	.	1	.	1	.	.	=	5
E4:	q_2	.	.	.	1	.	1	.	=	5
E5′:	q_3	.	.	.	1	−1	.	1	=	3

'e put the non-basis variables (x_{DS} and q_1) = 0, and obtain the
ion

$$\begin{aligned} x_{AR} &= 95 \\ x_{BS} &= 100 \\ x_{DR} &= 5 \\ q_2 &= 5 \\ q_3 &= 3 \end{aligned}$$

at a cost of $12 \times 95 + 8 \times 100 + 2 \times 5 + 0 \times 5 + 0 \times 3 = 1$

The reduction of £50 in the cost is as expected, as we have rep 5 tons of A by 5 tons of D at a net saving of £10 per ton repl The economic values of DS and q_1 are now

$$\mathcal{Z}_{DS} = c_{BS}.1 + c_{q2}.1 + c_{q3}.1 = 8$$
$$\mathcal{Z}_{q1} = c_{AR}. - 1 + c_{DR}.1 + c_{q3}. - 1 = -12 + 2 = -10$$

Hence $(\mathcal{Z} - c)_{DS} = 8 - 2 = 6$

and $(\mathcal{Z} - c)_{q1} = -10 - 0 = -10$

This value of -10 for $(\mathcal{Z} - c)_{q1}$ shows what we should lose removing 1 ton of D from R and replacing it with 1 ton of A.

We do not need to work out the values of $\mathcal{Z}$ and $(\mathcal{Z} - c)_j$ by method at each step. We can calculate these values by treating part as another equation as shown below, and this is often a sim method. The economic value of any activity in the basis is its cost, and hence $(\mathcal{Z} - c) = 0$.

Going back to Matrix II, the required equation is found by mu plying E1, E2, E3, E4, and E5 by the unit cost of AR, BS, q_1, q_2, an respectively and adding:

12.E1:	$12.x_{AR}$			+	$12.x_{DR}$			=	1 200
8.E2:		$8.x_{BS}$				+	$8.x_{DS}$	=	800
hence	$12.x_{AR}$	+	$8.x_{BS}$	+	$12.x_{DR}$	+	$8.x_{DS}$	=	2 000
deduct the functional	$12.x_{AR}$	+	$8.x_{BS}$	+	$2.x_{DR}$	+	$2.x_{DS}$	=	$\mathcal{Z}$

to obtain the new equation

$$EZ: 0.x_{AR} + 0.x_{BS} + 10.x_{DR} + 6.x_{DS} + 0.q_1 + 0.q_2 + 0.q_3 = 2\,000$$

and this equation would be written at the foot of Matrix II. 'factor' for the substitution is 10, the coefficient of x_{DR}. Multiply by this factor:

$$E3'\ (Z): 10.x_{DR} + 10.q_1 = 50$$

Deduct E3′ (Z) from EZ:

$$EZ': 0.x_{AR} + 0.x_{BS} + 0.x_{DR} + 6.x_{DS} - 10.q_1 + 0.q_2 + 0.q_3 = 1\,950$$

$(\mathcal{Z} - c)_{DS}$ is positive, so DS must go into the basis. By the division we substitute it for q_3. The coefficient of x_{DS} in E5′ (= 1) is the pi and the 1 in E2, the 1 in E4, and the 6 in EZ′ are the factors. Dividing by the pivot, we obtain

$$E5'': x_{DS} - q_1 + q_3 = 3$$

Multiplying by the factors, we have

$$E5''\ (2): x_{DS} - q_1 + q_3 = 3$$
$$E5''\ (4): x_{DS} - q_1 + q_3 = 3$$
$$E5''\ (Z): 6.x_{DS} - 6.q_1 + 6.q_3 = 18$$

leduction from E2, E4, and EZ′ we obtain

$E2'$: $x_{BS} + q_1 - q_3 = 97$

$E4'$: $q_1 + q_2 - q_3 = 2$

EZ'': $0.x_{AR} + 0.x_{BS} + 0.x_{DR} + 0.x_{DS} - 4.q_1 + 0.q_2 - 6.q_3$
$= 1\,932 - Z$

matrix is:

Matrix IV

		x_{AR}	x_{BS}	x_{DR}	x_{DS}	q_1	q_2	q_3	P
E1′:	x_{AR}	1	.	.	.	−1	.	.	= 95
E2′:	x_{BS}	.	1	.	.	1	.	−1	= 97
E3′:	x_{DR}	.	.	1	.	1	.	.	= 5
E4′:	q_2	.	.	.	.	1	1	−1	= 2
E5″:	x_{DS}	.	.	.	1	−1	.	1	= 3
EZ″:		0	0	0	0	−4	0	−6	

$= 1\,932 - Z$

his gives the best solution, as no coefficient in EZ″ is positive, and total cost is £1 932.

he above example points to one of the difficulties that arise in ar programming: the tendency for the matrix to become large (this mple has only two products and three ingredients, yet this has led system with five equations in seven unknowns). If all the items in matrix were 0's and 1's even a very large matrix could be solved by d, but if there are more than a few columns of decimals the time uired is disproportionately increased. Recent developments have, ever, shown that it is not always necessary to write a matrix in and certain limits can often be omitted until quite a late stage in calculations.[2]

n practical cases the optimal solution can sometimes be reached by rocess of inspection and trial and error (depending on the com- ity of the problem); but the use of linear programming ensures that optimal solution is eventually reached and is especially useful where prices fluctuate, as there will be critical values such that at these ts the nature of the optimal solution varies.

s an illustration of this point, suppose the price of D rises to £8 per equalling the price of B. At this stage the optimal solution reaches tical point, as if the price rises any further we shall not use D in S. n D at, say, £9 per ton, we find, from Matrix IV, that

$$Z_{q1} = (Z - c)_{q1} = -12 + 8 + 9 + 0 - 9 = -4 \text{ and that}$$
$$Z_{q3} = (Z - c)_{q3} = -8 - 0 + 9 = +1$$

should therefore put q_3 into the basis in place of x_{DS}, which would me a non-basis variable and equal to 0 in the new optimal solution. though the optimal solution is one of the most important results, by no means the only useful information given by the matrix. To

quote a few others: the cost of the restrictions can be shown; the
often several alternative solutions only very little more expensive
the optimum; the values of $(Z - c)$, being the difference be
economic values and actual prices, can be very useful to the buy

BIBLIOGRAPHY

[1] Charnes, A., Cooper, W. W. and Henderson, A. (1953). *An Introduction to Programming*. Wiley, New York; Chapman and Hall, London.

[2] Dantzig, G. B. (1955). 'Upper bounds, secondary constraints and block tri larity in linear programming', *Econometrica*, **23**, 174.

[3] Lomax, K. S. (1952). 'Allocation and programming in modern econo Manchester Statistical Society.

[4] Symposium on Linear Programming (1955). *J. R. Statist. Soc.*, B, **17**, 165.

[5] Williams, N. (1955). 'An application of linear programming to the selecti raw materials', *Applied Statistics*, **4**, 22.

INDUSTRIAL CENSUSES IN GREAT BRITAIN AND THE UNITED STATES

RITA J. MAURICE
University College, London

At the end of October 1954 the Verdon Smith Committee appointed to advise the President of the Board of Trade on future policy on Censuses of Production and Distribution made its report. Some six months earlier an assessment had been made by the Intensive Review Committee of the American statistical programme carried out by the Bureau of the Census. Miss Maurice compares the post-war policies of the two countries on industrial censuses, which were examined by these committees, and gives tables showing the kinds of information which have been collected.

round

e American Census of Manufactures has a much longer history its British counterpart. Corresponding to its longer history, pre-\merican enquiries were more frequent and more searching than made here. A full census was taken there every two years, as ed to once every five years in this country, and although both ses asked for a description of the establishment and enquired employment, raw materials, contract work, production, and inery, the American censuses also covered a number of other . Of these, wages were specifically excluded from the British s by the 1907 Census of Production Act. Detailed questions on ity and cost of materials and fuel were provided for only in the amendment to this Act which the war prevented from being put ffect. Inventories of finished products and materials, and capital diture on new plant and equipment, introduced respectively in merican 1937 and 1939 censuses, reflected the greater use of the can census for the provision of economic indicators. Both coun-sed the census as a basis for an index of the volume of production, 1935 the American report stated another purpose in taking the —besides showing production of various goods, relative size of ries, and industrial importance of areas—to be the measurement nthly and long-term trends in employment. For the census to be in this way speedy publication of results is required, and publica-f the 1939 American census was completed in 1942, whereas ation of the British 1935 census was not finished when war broke 1939.

ar Decisions

war interrupted census-taking in both countries and in each

case led to a review of the programme. In America, wartime was concentrated on the collection of special commodity series, were published monthly or less frequently in *Facts for Industry* r started in 1943. The content and coverage of these series was is) determined in consultation with government and industry use effort being made not to overlap with other government agenc trade associations. In 1955 there were fifty-four commodity s covering about one-third of the individual products covered i Census of Manufactures. It was decided after the war to take census of manufacturing once every five years instead of every year as previously, in order to save money on the collection of which was not considered necessary. Also, frequent changes in class tion, definition, and coverage had impaired the usefulness of mor quent censuses. This change was made possible by instituting an a survey of manufactures on a sample basis which, together wit commodity series, satisfied the government's and businessmen's for current data. The annual survey fills out the information avai from commodity series by providing general statistics of manufact activity for manufacturing as a whole and for important indiv industries and groups of industries, geographic divisions, and States full census is still the source of detailed relationships and of data for areas, and provides the necessary 'benchmark' on which to base the ple surveys. This use of sampling in the programme was an e new development in the manufacturing statistics of the United which was the first country to adopt this type of programme.

In Great Britain, the war caused the suspension of most short- statistics also, but with the end of the war a wider range of prod series than before was made available in the *Monthly Digest of St* which was begun in 1946. In 1955 there were forty-six tables of collected by various government departments and certain trade ciations covering commodities made by the more important facturing industries. Power to collect short-period figures was by the 1947 Statistics of Trade Act which also required an census to be taken of production. An annual enquiry had bee posed in the White Paper on Full Employment Policy approv Parliament in 1944. The Census of Production Committee, rep in 1945,[1] assumed this and recommended that powers be giv the collection of additional census information, the questions asked being decided each year in conjunction with an advisory mittee of experts from government service, research workers, businessmen. They also pointed out the need for supplementing c information by indicators of future activity, such as orders on and capital expenditure commitments, to be collected more frequ than once a year. The authorised programme was started in 194 was designed to provide:

'(1) Annual data on the structure of all industry and the bulk employment, inventories, and production and

) less frequent figures (possibly, once every three years) for larger s, on the use of resources, the details of gross output and industrial s, and on sales distributed by type of customer.'[2]

hus both countries established an elaborated version of the type rogramme which was suggested by the United Nations Statistical mission a few years later. That is to say:

1) infrequent detailed enquiries on structure, labour and some tal resources, and investment and production activities for the bulk ndustry, and

2) annual aggregates on employment, and investment and produc- activities for important industries.'[3]

he United States, however, sampling was an integral part of the gramme from its commencement. Nevertheless, although the ish 1947 Act required a census each year, it allowed for exemption such categories as the Board of Trade might determine', thus per- ing a sample enquiry although not requiring it. In America too, t-period series were an essential part of a co-ordinated programme, olicy of integration which was recommended in Britain in 1954 by Verdon Smith Censuses of Production and Distribution Com- tee.[4]

First Post-war Censuses

t Enquiries. Since neither country had conducted a census for so g, both departed from previous practice and preliminary tests were de before the full census was undertaken. In America four pretests e carried out solely to determine the ease of reporting new ques- s. As a result enquiries were included in the 1947 Census on oloyment and payrolls for one pay period, man-hours by quarters, capital expenditure in detail. Questions were not asked, however, containers used or purchased, on capital assets, on transportation s and weight of shipments, and on a breakdown of inventories of ds-in-progress. (It is interesting to note that all of these, except ght of shipments, have been included in British schedules and have caused particular difficulty.) From the enquiries it seemed that size of the establishment was more important in determining the e and quality of the replies than was the design of the form. In ticular, the size of the establishment was more important in obtain- replies to questions on man-hours, inventories, transportation, perty account, and capital expenditure than to questions on employ- nt, wages, and containers. In addition to the pretests, preliminary fts of product enquiries were sent to manufacturers and associations, ose comments and recommendations helped to determine the list products used in the census.

n Britain a partial census was carried out in 1946 of eleven indus- s chosen primarily on the basis of the need for information but also

in order to show up future problems. Questions were asked fo
first time on wages and salaries, and a number of definitions w
changed from those used before the war in order to make the fi
more consistent and the schedules easier to complete. Prelim
results of the partial census were published in trade reports in the *A*
of Trade Journal between March 1948 and March 1950. In the
of the results a number of amendments were made to the census f
issued for the first full census relating to 1948; but most of the cha
made in the partial census were retained (e.g. transfers betw
departments were adjusted, if necessary, to the value which w
have been charged to an independent purchaser; sales and purch
were not adjusted for stocks in detail).

Coverage. Special preparations had also to be made for the census
of establishments, as pre-war lists were out of date. In America
pre-war list was matched against the records of the Bureau of Old
and Survivor's Insurance to form the basis of the new list, which
supplemented by information from other agencies and from tr
directories. The current British list was the Factory Inspectors' regis
which is kept up to date by the periodic inspection of all factories us
machinery. The Ministry of Works keeps a list of building firms,
one of mines is available from the Ministry of Fuel and Power. No
is made here of insurance records, which have been found not sufficie
accurate, nor can Inland Revenue records be utilised, as these
confidential.

The American insurance records do not include establishments w
no paid employee and these were therefore excluded from the cen
instead of using the previous lower cut-off of $5 000 production.
Great Britain all manufacturers had to return some form, but th
employing ten or fewer in general stated only number employed
nature of output. In thirteen trades where small firms were relati
important these establishments returned a simplified form on sales
materials which varied from trade to trade. The similar Amer
short form was originally intended for only the smallest plants w
one employee, although it was eventually used by some larger
ployers. Thus practice was more uniform than it at first appears;
standard forms in America covered ninety-three per cent of employm
and in Britain the coverage was much the same. In both countri
preliminary mailing was used to determine whether establishme
were within the scope of the census and in America to find the nat
of the business, so that the correct schedule could be sent. The Uni
States Census Bureau used over 200 different forms as opposed to
British total of about 150. Coverage of the British census was
wider than that of the American, as it included building and contr
ing, public utilities, and mines and quarries, as well as manufactu
industry. Both excluded merchanting, but there were some differe
in the treatment of borderline cases (e.g. manufacturing in penal in
tutions and single-shop retail bakeries were included in Great Bri

not in America). There were also some changes from pre-war tice (e.g. the inclusion of textile jobbers in both America and at Britain).

dure and Response Rates. The methods of taking the census differed in the two countries. Both used a postal questionnaire on an blishments basis with generally separate returns for different lines ctivity at the same place. In America, thirty days were allowed he return of the schedule, and a supplementary mailing was made 0 000 out of 325 000 establishments circulated and 225 000 tabu-l. In the last six months of the year personal calls were made to ct schedules which had not been returned. British manufacturers allowed three months (as opposed to the two recommended by Nelson committee in 1945 and allowed in the partial census), and a postal follow-up was employed. Second reminders (following minder of the closing date) were needed for about 115 000 out of 000 establishments circulated and 193 500 tabulated. This con-in co-operation is also reflected in the fact that failure to return schedules resulted in only one court case in the United States, reas in Great Britain 125 prosecutions were made. The poor onse was attributed to the long period since the last census had been n in 1935, but even in America eight years had elapsed since the census. Nor have the figures improved since 1948, as may be seen the following table:

nsus ear	Registered Applications as Percentage of:		Prosecutions	Percentage of Larger Establishment Returns Defective
	Schedules Sent	Returns Received		
948	3·8	7·6	125	25
949	6·5	17·3	227	18
950	10·4	25·4	76	18
951	9·1	13·5	104	35

ures for 1948 and 1951 include short forms to smaller establishments; 1950 excludes ng and contracting.

eturns for the first full census related to the calendar years 1947 1948 respectively, but whereas American firms had to make a al application if they wished to make a return for their fiscal year, ritish firms were allowed, if they wished, to make returns for their ness year of account ending between 5th April 1948 and 6th April . Even when permission was granted, the American manufac-rs were required to give employment and payroll data on a calendar basis.

Questions Asked. The questions included in the respective censuses hown in the tables given at the end of this article. In the American census the classes of employees were consolidated from the 1939

detail of seven categories to three, but the man-hours enquiry wider coverage than previously. Although the materials shown s rately decreased from 275 to 150, they were more carefully chose the coverage of their use was improved. The product classifi resulted in published figures for about 6 100 commodities as com with 4 000 in 1939. Excise taxes were for the first time excluded product values as well as net output, and shown separately where were of importance. The capital expenditures enquiry was detailed than before, but there was no enquiry on power equip (included before the war) on the grounds that such slow-mo inventories need be measured only at ten-year intervals. The ques on motor vehicles used were new, and there was no analysis of s as the Census of Distribution was not taken until a year later.

The British schedule also showed considerable changes as comp with previous censuses. The separate listing of working proprieto other than limited companies had been recommended by the Census of Production Committee in order to relate the new informa on wages and salaries (which they had also recommended) to appropriate number of workers. An enquiry into man-hours was m in fifteen industries but has not been repeated since. Such informa proved difficult to collect and, as there are some data from Min of Labour enquiries, the Verdon Smith Committee recommended the questions should not be asked again unless there is a specific in relation to other census questions. Another recommendatio lowed only at this census was the enquiry into services rendere other firms (e.g. research, advertising) in order to make more mea ful comparisons of net output. Excise tax information was also lected in order to render this calculation more accurate. Other topics which had been recommended in 1945 were inventories, ca assets and expenditure, and sales by class of purchaser. Some enq suggested in 1945 were not included in 1948, however, and hav been started since then. The most striking of these concerns the co of establishments; it is necessary for any analysis of the concentr of production. In addition, no information has been collected o tory space and age of buildings, which are of interest for locati industry and town planning policies.

When comparison is made between the American and B schedules, the absence of a British enquiry into the control of fir again notable. This subject has been covered in all detailed Amer censuses since 1900. The British breakdown of the employment fig was on this occasion more detailed than the American, although largest category of 'operatives' is almost the same as the Amer 'production and related workers'. The British separation of in tories of materials and work in progress was not attempted in Am in 1947 on the basis of pretest results, but American practice has lowed the British since 1953. The new questions on capital a asked in Great Britain had no counterpart on the American sche

tal assets were last included in the American census in 1919 and, 947 pretests had shown that only about forty per cent could er such questions satisfactorily, had not been reintroduced. Nor they been included in subsequent British censuses.

cation. In the analysis of the American census a Standard Indus- Classification set up in 1940 was introduced, covering 450 indus- in twenty major groups. Some modifications were made, how- to prevent overlapping. Another move towards uniformity een various government statistics was made in the area tabulations the replacement of industrial areas by fifty-three Standard Metro- an Areas, used also for business and population figures. Similarly, the British 1948 census the Standard Industrial Classification h was drawn up in 1946, containing twenty-four groups under headings, was used, and the country was divided into the eleven dard Regions generally used for regional organisation of govern- t departments.

eliminary releases of the 1947 American census results began in mber 1948, and the four reference volumes (General Summary, stics by Industry, Statistics by State, and Product Supplement) published in 1950. The British performance contrasts unfavour- with this. Preliminary 1948 results were published in the *Board rade Journal* between April 1950 and July 1951, and the 156 final e reports were completed only in March 1953. The delay was buted to lack of co-operation by manufacturers and inexperience ne Census Office. Another important factor was the concession ntioned above) to manufacturers wishing to make a report for business year. The Verdon Smith Committee has now recom- ded a change towards American practice by requiring output es for the calendar year, but the period allowed for returning the remains two months longer than in the United States.

s of Accuracy. The 1947 American census was followed by an esting innovation in method which was first used in connection the 1945 Census of Agriculture.* This was a test of the quality e census, in this case to estimate primarily the completeness of rage. The sixty-eight sample areas employed in estimating figures urrent population and labour force were used, with additional pling of places of more than about 25 000 population and of factory s of smaller towns in order to improve industrial coverage. In 10 000 small sample areas chosen, all recognisable commercial blishments were listed and representatives of all those not on the us register were interviewed. The resulting list was checked against ensus register, which was itself compared with the list of establish- s included in the census tabulations. In doubtful cases further iry was made directly from the establishment or from the records

* For a summary of the quality checks undertaken by the US Census Bureau see 'Measuring the Accuracy of Enumerative Surveys', by A. Ross Eckler and Leon Pritzker. *Bull. Internat. Statist. Inst.*, **33**, part 4, 1951.

of the Bureau of Old Age and Survivor's Insurance. The r
showed that the census included about ninety per cent of the esta
ments and ninety-eight per cent of the employment and payroll
out to cover. The omitted establishments were thus smaller on av
(twelve employees) than those covered by the census (sixty emplo
Nearly half of the omissions were not covered by the register,
the remainder were sent schedules but did not return them, or
incorrectly classified as non-manufacturing or out of business or
basis of their return. Very few of the largest establishments
omitted, coverage of those with over fifty employees being over nin
eight per cent, and the results may therefore be considered
satisfactory.

Subsequent Censuses

1949. After 1947, the next enquiry made in America was the
Annual Survey of Manufactures for 1949. The coverage was m
the same as that of the 1947 census, the sample of 45 000 establishm
(from some 260 000–270 000) including about two-thirds of t
manufacturing employment. It was selected on a company basis
stratified by industry group and size of establishment. All estab
ments with 250 or more employees were included. Below this
independent establishments were divided into six strata in which b
sampling rates varied from fifty per cent for the largest establishm
to two and one per cent for those employing less than ten. The car
was almost completely by post, personal visits being made in or
few very special cases. This procedure has proved satisfactory and
been followed for all subsequent surveys. The schedule used
common to all industries (with one or two exceptions) and w
shortened form of the general one used in 1947. Instead of sta
production of individual commodities, figures were asked for by '
duct class', but figures were published for only 109 out of the 1
classes. There was a new question on products bought and r
without manufacture, and also a new section comprising a check
of five questions by means of which the respondent could check
completeness and consistency of his own figures (e.g. by calcula
the average hourly wages implied by the wages and man-hour figu
This innovation proved successful and saved the Census Bureau m
trouble in editing and following up faulty returns, and it has b
included in all of the subsequent census schedules. Classification
industries in 1949 was as usual on the basis of the principal prod
but in certain cases of this criterion leading to a changed classifica
because of a change in the emphasis of production equal weight
given to the current and new status. This procedure led to gre
comparability of general figures between 1947 and 1949–53.
general figures for 1949 were obtained by using difference estim
from 1947 values for the sample, while for detailed statistics (such
electric energy purchased) an inflated estimate from 1949 data al

used, since this method of estimation was cheaper. Approximate dard errors were calculated for the estimates and rounded up (to nearest whole number if under 5% and to the nearest 5% if larger this) before publication. For most United States and industry p totals these were under 5%, although for a few State and small stry figures the errors were as high as 30–40%.

Great Britain the full census of 1948 was followed by further com- coverage enquiries, although the information collected was more ss restricted. In 1949 the questions were simplified and the shorter to smaller establishments was dropped. Firms reported only total uct, and this meant that unless they reported a substantial change eir output they were classified in the same way as in the previous This corresponded with the American procedure and gave ter comparability between years, although the check of accuracy ded by comparing product and materials data was lost. There a supplementary voluntary enquiry into capital expenditure in year by a sample of nearly 4 000 establishments in order to obtain nformation more quickly than was possible from census returns. ough the sampling was stratified by industry and size, some of the pling errors were large (e.g. 23% for the clothing industry) but the ltaneous collection of information for 1948 enabled more accurate ates of change than of level. This was the beginning of the use mpling in British industrial enquiries.

The American Annual Survey for 1950 corresponded exactly that for 1949 apart from a few small changes of definition. The ple, however, was modified to correspond with individual industry fication rather than the industry groups used in 1949. There was a marked improvement in the speed of publication, the first results g issued in November 1951, only three months after those for 1949. ne coverage of the British census was changed to exclude building contracting firms, and firms in trades combining production and ing activities, in order to avoid overlapping with the Census of ibution. To supplement information from this census questions asked for the first time on merchanting activity and there was nalysis of sales. This had not been part of an American census 1939. However, since these questions proved difficult to answer exclusion from future censuses is to be considered on the basis ese results. There was another sample enquiry into capital expen- e, the sample being modified on the basis of the previous year's ience to include relatively more larger establishments.

For 1951 the basis of the American sample design remained the but the sample was again modified in order further to reduce mpling vacancies and improve product class figures. The most rtant change on this occasion was the use of the UNIVAC elec- computer for production statistics for the first time. Since then s been used for each of the surveys and in 1954 was used for

editing and coding as well as tabulating the returns. Its use m possible more detailed tabulations and speedy publication of re Improved sample design, better processing, and greater confider the results after three years' experience led to the publication of fi for 652 product classes, and this helped to meet some of the criti which had been made of the Annual Surveys. A new and special was distributed to all of the largest and a sample of the smaller panies enquiring into capital expenditure on new or renovated e lishments and separate tables were published for this form investment.

Coverage of the British census was restored by the reinclusio building and contracting. The schedules were more detailed than had been since 1948 and the short form was again sent to cer smaller establishments. The question on transport payments modified so that there was no need to analyse transport account order to make the return—yet another change designed to make schedule easier to complete. The census capital expenditure quest were restored and no sample enquiry was made. As in the Un States, information was collected for the first time on capital expe ture at establishments not yet in production.

1952. There were no changes in the procedures or questions of American survey, but previous performance was improved u Figures were published for 686 product classes and the results b to be issued in August 1953 and were completed by March 1954.

There was a fundamental change in the British census, how with the introduction of sampling for the main enquiry. This foll the practice which was part of the original American plan, the ch being made in order to obtain earlier estimates. Although the re of the 1949 and 1950 censuses were completed more quickly than for 1948 (by March 1953 and October 1953 respectively) the imp ment was not great in view of the greater simplicity of these cen Since greater output and material data (considered too detailed collected by sampling methods) had been collected in 1951, a sa census was sufficient to collect the summary figures required. largest firms were all included in the sample, the size criterion fo approximating to the American one of 250 or more employees. each trade smaller firms were divided into two or three strata. sampling rate for those employing more than ten was approxima one in three and for those employing ten or less it was about on twenty. The 1952 schedule was again more simple and fewer quest were asked than in any previous census.

1953. Extensive preparations had been made and some adva notices were sent out for another American census of manufact due in 1953 as part of the quinquennial census programme for ma facturing, business, mineral industries, and transport. However, C gress failed to make the necessary appropriations, although money

ided for 'spot checks'. From this appropriation another sample
ey on the same pattern as those in previous years was financed.
his occasion a sample of about 50 000 establishments was newly
n on the basis of employment records of the Bureau of Old Age
Survivor's Insurance. The schedule included a new enquiry on
strial water intake and more complete and searching check ques-
. Publication again began in the following August but later
rts were held up by preparations for the full census relating to
which was approved in that month.

ne British 1953 census was again a sample and simplified enquiry
the object of speeding up the returns. For the processing of this
as the American lead was followed (although on a much smaller
) and use was made of an electronic calculator. The 'investigation
e use of the new forms of electronic machinery' was recommended
ne Verdon Smith Committee, and the calculator was installed
in 1955. As a result, procedures have been speeded up and the
has been reduced by about one-fifth from the April 1954 size of
The first results for the 1953 census appeared in July 1955, before
for 1952. Thus, as in the United States, the advantages of a
le enquiry were not realised until some experience had been gained.
ampling errors were published with these figures.

The full American census was patterned largely on the basis of
ous censuses, and therefore the pretest programme was not so
sive as for 1947. However, between 1950 and 1953 several special
es were made on the reporting of materials consumption, as
er detail was required for certain critical materials and in order
ake possible inter-industry studies of the flow of materials. At
census a new summary form for the company was issued asking
summary of reporting units and employment, including employ-
falling outside the scope of the censuses. An enquiry into power
oment was included for the first time since 1939. To ascertain
mportance of the establishments with no employees which were
ed from the census as in 1947, a sample was taken from the
nal Revenue Service Records. Advance reports for a few indus-
had been published by the end of 1955, but no summary results
been issued.

full census of all establishments was also taken in Great Britain.
new enquiry was made on employers' contributions to superan-
on funds with the object of making more complete the information
bour costs. The questions corresponded in detail with those of
apart from materials, for which the greater 1948 detail was
cted. The electronic calculator was on this occasion used for
ng the returns, as in the United States. Preliminary summary
s (based on about seventy per cent of the returns) were published
e end of 1955, only five months after those for the 1953 sample
ry.

Thus the British census programme has recently been nearer
American than at its inception. Advantage has been taken
economy and speed of sampling when the detail required fro
census has allowed. This type of programme is to be continu
future, as the committee which considered future policy recomm
a triennial census (with the possible lengthening of the interval t
years after 1960) and sample surveys for the intervening years.
recommendation was made on the grounds of the need for relia
in the provision of a sampling frame, and in the reflection of pr
changing economic conditions. However, the Board of Trade
to allow four years to elapse before taking the next full census in
The committee also recommended the use of sampling for ce
topics in the full census and the raising of the exemption limit
eleven to twenty-five employees. This latter recommendation
further move away from American practice, but sampling of
smaller firms from time to time was suggested in order to be ab
estimate the total output of each trade. The most important o
other post-war developments in method made in America has als
cently been put into effect here, with the use of an electronic calcul
It is hoped that in 1956 the US Census Bureau will extend the u
its computer to the analysis of commodity series. In view of the
in publication that has hitherto been experienced in this country
to be hoped that this development will result in more timely re
Introduction of the American innovation of check questions w
also help considerably here in view of the high percentage of re
received which are defective. The other development which
introduced first in America is the check of the quality of the
sus. This is perhaps of more academic interest than the other
but a favourable quality check would no doubt help to mee
lack of faith in the reliability of the census which has been
by business organisations as a reason for not making greater u
census data.

Value of the Census

It is in the extent to which census data were found to be use
businessmen that the reports of the American Intensive Review
mittee[5] and the British Verdon Smith Committee[4] show the gr
contrast. The governments of both countries use the census as a b
ground for economic policy decisions, as a 'benchmark' for short-pe
figures, and as the basis for an index of industrial production. Ce
information is also used in preparing national income estimates
means of a 'mark-up' to estimate total expenditure in the United St
and for wages and salaries, capital expenditure, stocks, and s
classes of personal expenditure in the United Kingdom. It also
vides data needed for productivity studies and input/output anal
The necessity of the census for defence planning was also recog
by both the American and British census committees. In both cou

he census is used by trade unions in the formulation of their es and by research workers.

e OEEC Technical Assistance Mission found that the problem in Jnited States was to pare down demands for information rather to strive for business and public recognition.[6] This is supported e examples of firms financing the continuation of tabulations appear to be no longer warranted by the 'public interest', or pay- r special tabulations (e.g. to compare their own performance with ggregate). Such work is encouraged by the US Census Bureau, s extension is now recommended in this country. In the survey d out by the American Marketing Association for the Intensive w Committee, of the 35% of its members who replied, three- ers used the Census of Manufactures, its use being in extent and ncy second only to that of the Population Census; 50% used it les forecasting; 62% in analysing market potentials; 35% in ng sales performance and in analysing distribution; and 30% in yout of sales territories. Only 13% wanted the full census less ntly and 38% wanted it more frequently than once in five years, ample enquiries are not sufficiently detailed or reliable for many ses. Although the replies were biased in favour of companies ying research staffs, they covered a wide range of activity and st. The US Chamber of Commerce carried out a similar survey of associations with similar results. These surveys produced various tions for improvement of the programme and the Intensive w Committee recommended greater detail by materials used and lustrial processes, with separate reports for different products at me establishment. They also took up the requests made for r speed in publication and emphasised the need for more timely —then more than a year ahead of the British!

ontrast, the British committee found that trade and commercial sations and firms make little use of census information. This is ted to delay in publication, failure to publish sufficiently detailed ation, and lack of comparability with other figures casting on the reliability of the data. In line with this and the fact e committee estimated that the cost of a census is greater to y than to the government (this has been estimated also for eriod statistics in America) it was considered that the census be no more frequent than was 'undeniably necessary'. Some es suggested that at least five years should elapse between s and the National Union of Manufacturers suggested an even exemption limit of one hundred employees. The committee put for- or consideration further simplifications to the census forms, and nsidered that before any new questions were introduced the Board e should make certain of the need for them and that the census was way to obtain the information. Thus the emphasis was on limit- enquiries in order not to place too heavy a burden on industry as l to the American recommendation of greater detail.

TABLE A

Details of manufacturing enquiries in the United States, 1947–54

	1947	1949	1950	1951	1952	1953
Name of establishment	×	—	—	—	—	—
Location of establishment	×	×	×	×	×	×
—within corporate boundaries	×	×	×	—	—	—
Social Security identification number	×	—	—	×	×	×
Change of owner in census year	×	×	×	×	×	×
Change of working status in census year	—	×	×	×	×	×
Type of ownership	×	—	—	—	—	—
Controlling or controlled by another company	×	—	—	—	—	—
Period to which return relates	×	×	×	×	×	×
Nature of activity of establishment	—	—	—	—	—	—
Employment in particular pay periods:						
total	12, 1a	4	4	4	4	4
production and related workers	12, 1a	4	4	4	4	4
force account construction workers	1a	} 4	4	4	4	} 4
clerical, routine office workers, etc.	1a			} 4	4	
all other workers	1a					
Total man-hours, by quarter, of production and related workers	×	×	×	×	×	×
Wages and salaries paid in year:						
production and related workers	×c	×	×	×	×	×
force account construction workers	×c	} ×	×	×	×	} ×
all other workers (administrative, supervisory, etc.)	×d			×	×	
Cost of materials, fuel, and electric energy:						
materials, parts, containers, and supplies	×	×	×	×	×	×
products bought and resold without manufacture	—	×	×	×	×	×
fuel	×	×	×	×	×	×
purchased electric energy (cost and amount)	×	×	×	×	×	×e
contract work	×	×	×	×	×	×
Quantity and cost of number of fuels	6	—	—	—	—	—
Electric energy generated	×	—	—	—	—	—
Electric energy sold	×	—	—	—	—	—
Materials consumed—individual detail	×	—	—	—	—	—
Industrial water intake	—	—	—	—	—	×
Value of products shipped:						
total	×	×	×	×f	×f	×f
individual detail	×	—	—	—	—	—
product class detail	—	×	×	×f	×f	×f
Amount received for:						
commission work	×	×	×	×f	×f	×f
scrap and refuse	×	×	×	×f	×f	×f
repair work	×	×	×	×f	×f	×f
installation and construction own products	—	×	×	×f	×f	×f
research and development	—	×	×	×f	×f	×f
resale without manufacture	—	×	×	×f	×f	×f

TABLE A (*continued*)

	1947	1949	1950	1951	1952	1953	1954
l products producers enquiry:							
etals consumed	×	×	×	—	—	×	—
erations performed	×	—	—	—	—	—	—
ntories at beginning and end of :							
ished goods	×	×	×	×	×	×	×
rk in progress	×	×	×	×	×	×	×
aterials, fuel, etc.	×	×	×	×	×	×	×
nditure for plant and pment:							
w construction—building ..	×	×	×	×	×	×	×
—other ..	×						
w machinery and equipment —production	×	×	×	×	×	×	×
—other (office machinery, vehicles, etc.)	×						
nd	×	—	—	—	—	—	—
her (used plant and machinery, etc.)	×	×	×	×	×	×	×
ital expenditure at establish-ents not yet in production	—	—	—	×	×	×	×
e-power rating of:							
ime movers and generators ..	—	—	—	—	—	—	×
ectric motors	—	—	—	—	—	—	×
way-type motor vehicles erated (by types)	×	—	—	—	—	—	—
cks for completeness and curacy	—	×	×	×	×	×	×

nbers corresponding to employment figures indicate number of figures given for the f enquiry. Other numbers indicate number of subdivisions.
Distinguishes male and female.
Twelve for some industries.
Figure also given for one particular pay period.
Figure also given for clerical, routine office workers, etc., for one particular pay period.
Quantity relates to total, whether purchased or not.
Figure for the previous year also given.

ne American Census Bureau makes use of a number of advisory nittees, some of which are appointed specially to consider a census e others, such as the Advisory Committee on Federal Reports h advises on the total government statistical programme, are on ore continuous basis. It consults with some six hundred trade iations which provide representatives for these committees and generally co-operative. The Board of Trade also consults with t five hundred trade associations, but the Verdon Smith Com-ee found various specific difficulties in completing trade forms, h suggests that the closer contact of the Board of Trade with stry and trade which they recommended, and which is now being e, is particularly necessary. The committee also emphasised the to improve public relations by publicising the usefulness of the us and as a means of securing greater co-operation, which would e possible the publication of more useful results and provide a basis for the future programme.

TABLE B

Details of production enquiries in Great Britain, 1948–54

	1948	1949	1950	1951	1952	1953
Title of firm	×	×	×	×	×	×
Address	×	×	×	×	×	×
Address of registered office if limited company	×	×	×	×	×	×
Change of owner in census year	—	—	×	×	×	×
Period to which return relates	×	×	×	×	×	×
Nature of activity	×	×	×	×	×	×
Change in nature of activity in year	—	×	×	—	×	×
Number of establishments to which return relates, with details of each	×	×	×	×	×	×
Working proprietors	×a	×a	×a	×a	×	×
Employment in particular pay periods:						
operatives	12, 1ab	1ab	1ab	1ab	—	—
administrative, technical, clerical workers	1ab	1ab	1ab	1ab	—	—
canteen workers	1a	1a	1	1	—	—
all other workers	1a	1a	1	1	—	—
Selling staff	1	—	—	—	—	—
Transport workers	1	—	—	—	—	—
Workers on maintenance of buildings	1	—	—	—	—	—
Research staff	1	—	—	—	—	—
Outworkers	4a	—	—	—	—	—
Average number employed in year:						
operatives	×	—	×	×	×	×
administrative, technical, clerical workers	×	—	×	×	×	×
outworkers	×a	—	×a	×a	—	—
Shift working by operatives	—	—	—	×	—	—
Wages and salaries paid in year:						
operatives	×c	×c	×	×	×	×
administrative, technical, clerical workers	×c	×	×	×	×	×
outworkers	×	×	×	×	—	—
Employer's contribution to National Insurance	×	×	×	×	—	—
Employer's contribution to pension funds	—	—	—	—	—	—
Costs of materials, fuel, and electric energy:						
materials, parts, and supplies	×					
containers	×					
fuel	×	××	×	×	×	×
purchased electric energy	×					
contract work	×	×	×	×	×	×
Quantity and cost of number of fuels	6	—	—	4d	—	—
Electric energy generated	×	—	—	×e	—	—
Electric energy sold	×	—	—	×	—	—
Materials consumed—individual detail	×	—	—	×	—	—
Payments made for services rendered by other firms (repair, research, etc.)	×	—	—	—	—	—

TABLE B (*continued*)

	1948	1949	1950	1951	1952	1953	1954
ue of products sold:							
dividual detail	×	—	—	×	—	—	—
tal	×			×			×
aste products	×		×	×			×
ount charged for:		×			×	×	
mmission work	×		×	×			×
pair work	×						
ue of goods resold without anufacture	—	—	×	—	—	—	—
ounts paid for transport ..	×	×	×	×	×	×	×
ntories at beginning and end of :							
nished goods	×	×	×	×	×	×	×
ork in progress	×	×	×	×	×	×	×
aterials, fuel, etc.	×	×	×	×	×	×	×
enditure for plant and pment:							
w plant and machinery ..	×	×	—	×	×f	×f	×f
ed plant and machinery ..	×f	×f	—	×f			
w vehicles	×	×	—	×	×f	×f	×f
ed vehicles	×f	×f	—	×f			
red plant and machinery ..	×	—	—	—	—	—	—
w buildings and construction	×	—	—	—	—	—	—
d buildings	×f	—	—	—	—	—	—
nd	×f	—	—	—	—	—	—
ue buildings and land beginning year	×	—	—	—	—	—	—
ital expenditure at establishents not yet in production	—	—	—	×	×	×	×
se-power rating of:							
ime movers and generators ..	—	—	—	7g	—	—	—
ectric motors	—	—	—	7g	—	—	—
lysis of sales by methods of disosal	13	—	10	—	—	—	—

mbers corresponding to employment figures indicate number of figures given for the of enquiry. Other numbers indicate number of subdivisions.

Distinguishes male and female.
Distinguishes workers under eighteen.
Figures also given for each quarter.
Quantity only distinguishing four uses.
Electric energy used also given.
Figures for sales also given.
In reserve or idle distinguished.

ne British census is in one respect, however, more securely based the American, since money does not have to be specially appro- ed each year. This was the reason for the postponement of the rican full census from 1953 to 1954, and throughout the funds for Annual Surveys have not been part of the regular budget of the us Bureau, although it is hoped that they will be so in future. The of the American Annual Surveys is about $300 000, whereas the of a census is about $3 000 000. For 1950–53 British costs averaged

£180 000 a year, two of these enquiries having been on a sample b
The far greater cost of full census-taking in America thus well ju
the emphasis there on sample enquiries, and the economy in B
partially accounts for the delay in publication, although it is op
question whether this is in fact not a false economy.

Acknowledgement

I should like to express my thanks to members of the US Bure
the Census and of the Statistics Division of the Board of Trad
providing me with information.

REFERENCES AND BIBLIOGRAPHY

[1] *Report of the Census of Production Committee* (1945). Cmd. 6687. HMSO, Lond
[2] *Studies in Methods* (1953). Industrial Censuses and Related Enquiries. Volu
Text. Provisional Issue. United Nations.
[3] *Basic Industrial Statistics* (1951). E/CN. 3/124. Add. 1, 14th May, 1951. U
Nations Economic and Social Council, Statistical Commission.
[4] *Report of the Committee on the Censuses of Production and Distribution* (1954). Cmd.
HMSO, London.
[5] *Report and Exhibits to the Report of the Intensive Review Committee to the Secreta
Commerce* (1954). Appraisal of Census Programmes. United States De
ment of Commerce.
[6] *Industrial Censuses in the United States* (1955). Technical Assistance Missio
Organisation for European Economic Co-operation.

UNITED STATES: Reports of Censuses of Manufactures and Annual Surve
Manufactures.

GREAT BRITAIN: Reports of Censuses of Production and Board of Trade Journ

1950 *Britannica Book of the Year*. Census data, US 1947. Census of Manufa
Encyclopaedia Britannica Inc.

Industrial Censuses in Western Europe (1951). Report of a Group of Experts. Org
tion for European Economic Co-operation.

Some Informal Papers on the Industrial Statistics of the United States (1951). United
Department of Commerce, Bureau of the Census.

GEORGE, C. O. G. (1954). 'Industrial censuses in the United States', *The
porated Statistician*, **5**, 59.

LEAK, H. (1949). 'Statistics of the censuses of production and distribution',
Statist. Soc., A, **112**, 67.

A TECHNIQUE FOR STUDYING THE EFFECTS OF A TELEVISION BROADCAST

WILLIAM A. BELSON
Research Psychologist, Audience Research Department, BBC

This article describes a technique, *the stable correlate method*, for measuring the effects of television broadcasts seen under completely normal conditions. Mr Belson suggests that the method may be readily applicable to various sociological studies of effects, e.g. of television, of cinema-going, of major events, and of publicity campaigns.

duction

the summer of 1953 the BBC's Television Service broadcast a s of four programmes called 'Bon Voyage', the aims of the series g (i) to teach viewers French words and phrases of a kind likely to f use to them in making a first trip to France; (ii) to provide them general information for the same purpose; (iii) to remove appre-ions about possible language difficulties. An intensive study was e of the extent to which these aims were achieved, and the general ngs are reported elsewhere.[1] This paper, however, is concerned y with the *research technique* used in that inquiry. Whereas the tech-e was used specifically to assess the effects of the 'Bon Voyage' s, and is described in that context, it appears to have value as a eral research tool for the study of effects.

research design often advocated and used with advantage in studies ffects requires that people be tested before and after exposure to influencing conditions.[5] The assessment of the effects of a tele-n broadcast is, however, beset by difficulties which preclude the of a simple *before-and-after* design. In the first place, it is essential the viewing upon which the results are to be based should be ely normal (i.e. not influenced either by pre-broadcast testing or ny knowledge that tests are subsequently to occur). Secondly, it ten impossible with a television broadcast to say in advance of the ramme precisely what its content will be—as would be necessary ny effective *before-and-after* comparison.

f course, a superficial way out of these two difficulties would have to conduct *all* the test work *after* the series had been broadcast, method being to compare the test scores of those people who *did* he programme with those of non-viewers or persons who did *not* t. The difficulty here, however, was that the difference between two groups could not have been interpreted as pure effect, but as a mixture of effect and of original (or pre-broadcast) difference

between the two groups of people. After all, one of the groups el to view the programme while the other did not, and (apart from who might have wished to view but could not) this presupposes difference in attitude towards the subject of the programme. Ac ingly, for the *after-only* design to be of value, methods had to be f for partialling out or eliminating all or most of that part of the differ between test scores which was attributable to the fact that the gr were different to start with. Such a method was in fact devised was used with considerable success in the 'Bon Voyage' study. (method it had its roots in a form of partial correlation which I used earlier for reducing volunteer bias in invited groups and general matching purposes.[2] Wilkins[6] has developed techniques related kind in the field of criminological prediction.)

The Rationale of the Method

Basically, the method used was an adaptation of the more us matching technique, though unlike ordinary matching methods there was no discarding of subjects, (ii) the matching criteria w established empirically, (iii) it was possible to make a relatively f estimate of the success of the method in eliminating extran differences.

The rationale of the method may be summed up in the follow argument. Suppose that the non-viewers' knowledge of the broad French words is predictable (with a known degree of accuracy) thro a particular set of variables *which are not themselves open to influence b programme*. Suppose also that the viewers and the non-viewers d in respect of these prediction variables. Then, using a regression e tion based upon the prediction variables, it is possible to estimate w the non-viewers' score on the broadcast French words *would* have l had they equalled the viewers in respect of the prediction varia The residual difference between the scores of the two groups approximates to 'effects' of the programme. Clearly the higher multiple correlation between the prediction variables and the vari being studied (e.g. knowledge of the broadcast words), the great the confidence that the residual difference is 'pure effect' (and partly a result of extraneous differences). Accordingly the idea using the technique must be to build up the multiple correlation t high a level as possible. Strictly speaking, a multiple correlation anything less than $\pm$ 1·0 leaves room for doubt about the interpreta of residual differences, for it is still possible that the remaining varia springs from the operation of variables in respect of which the view and the non-viewers differ. It is therefore most important in us this technique to incorporate into the design a means of makin check upon the degree to which the chosen prediction variables h done their job.

The use of this approach in the 'Bon Voyage' study thus requi the development of both high-powered predictors and an effec

king device; some details of these are set out in the next two sec-
. Since it is a vital feature of the technique that the prediction
natching variables should not themselves be open to influence by
programme, they are referred to hereafter as the 'stable correlates'
e variable under study, and the method itself may for convenience
alled the 'stable correlate' method.

loping the Stable Correlates

s is typical of the method, no additional test-sessions were required
he development of the stable correlates. It was necessary in the
place to postulate a number of variables which it was reasonable
ink were associated with (and therefore predictive of) the variable
er study. For example, 'number of trips made to France' was
cted to be predictive to some extent of knowledge of the French
ds broadcast. Measures in terms of some six of these proposed
elates were then made *during the post-broadcast test sessions* (along
the main tests). After this, a preliminary analysis was made to
rmine, statistically, precisely what combination of the proposed
le correlates was most predictive of test score. Since in this par-
ar project three separate variables were under study, three separate
yses had of course to be made, the possibility being that three
rent combinations of correlates would be needed. It was essential
strict these analyses to the test results of the non-viewers because
nclusion of those of the other group would, if the programme had
ct produced changes, have led to a misleading attenuation of the
elations. The calculation was made through the Wherry-Doolittle
ula[4] and, as is usually the case, two to three of the proposed stable

TABLE I

The proposed matching criteria

ariable under Study	Stable Correlates Selected (in Order of Priority)		Multiple Correlation Achieved
	Variable	Correlation	
wledge of French	Control words	+ 0·78	+ 0·82 (first two)
ords and phrases pre-	Educational level	+ 0·59	+ 0·85 (all three)
nted in 'Bon Voyage'	Occupational level	+ 0·58	
wledge of facts pre-	Control words	+ 0·54	+ 0·63 (first two)
nted in 'Bon Voyage'	Visit to France*	+ 0·46	+ 0·65 (first three)
	Educational level	+ 0·51	+ 0·65 (all four)
	Occupational level	+ 0·43	
tude on issues related	Control words	+ 0·48	+ 0·52 (first two)
visiting France	Visit to France	+ 0·34	+ 0·55 (all three)†
	Occupational level	+ 0·39	

hose who had made a visit in the past were more favourably and confidently disposed
king a further visit.

o have selected more correlates than those shown would, in the case of each of the
les under study, have involved no appreciable increase in the multiple association
Wherry-Doolittle formula).

correlates taken in combination came very close to maximising co- tion. Thus in Table I, which sets out the main results of this liminary analysis, the two main predictors of 'knowledge of the b cast French words' yielded a multiple correlation of + 0·82, and figure would have been raised only slightly (+ 0·85) by the add of a third stable correlate.

This made it possible to develop, for each of the three vari under study, a short regression equation which, as already stated, to be used to adjust the score of the non-viewers. Such a corre formula, based upon two variables x_1 and x_2, would take the form

$$\tilde{y} = \beta_1 x_1 + \beta_2 x_2$$

(where $\tilde{y}$ is the corrected score and β_1, β_2 are constants, all the varia being measured from their means), although, as in the present c whenever the various distributions are not strictly normal or associations not strictly rectilinear, a simple weighting system as cribed below is preferable. Take for example the case of 'knowledge of broadcast French words', where the matching criteria are 'con words' and 'educational level'. If each of these variables had allo a division of subjects into three groups, then it would have been sible to classify the 120 subjects of the non-viewing group into sub-groups (e.g. control words X, educational level A; control w Y, educational level B; and so on), and these nine groups would have been the relevant ones as far as matching was concerned. aim of the weighting process would then be to determine what score of the total non-viewing group *would* have been if each of its sub-groups had contained the same number of people as were in comparable sub-groups of viewers.

TABLE II

Changes produced (post-broadcast groups)

Variable under Study	Non-viewers' Scores		Viewers' Scores	*P*
	Unadjusted Averages	Adjusted Averages	Unadjusted Averages	
Knowledge of words presented in 'Bon Voyage' (creative response)*	4·89	5·31	6·21	0·02
Knowledge of facts presented in 'Bon Voyage'†	8·54	9·08	10·10	0·02
Attitude on issues related to visiting France‡	15·38	17·65	11·80	0·01

* Score out of 26. † Score out of 29.

‡ 17·65 represents a relatively favourable and confident attitude.

NOTE. These figures are unadjusted for 'attendance bias' (i.e. less than half of the nu invited actually attended); where such adjustments *are* made, however, the estimat amount of change produced are virtually unaffected.

he effect of this correction, made through the weighting method, ustrated in Table II. Thus the unadjusted average score of non-ers on the word knowledge test was 4·89 out of the total of 26; on stment through the appropriate correlates this became 5·31, which compared with the average score of 6·21 by the viewers—indicat-that the programme had produced an increase (significant at the level) in knowledge of the broadcast words. To illustrate further, adjustment to the non-viewers' attitude score (through *its* corre-) moved that average from 15·38 to the more 'favourable' average 7·65, which then compared with the viewers' average of 11·80—cating that the programme had produced a sharp decrease in ers' confidence about going to France (significant at the 0·01). From this and the remaining data in Table II it will also be that in two of the adjustments the result was a reduction in the arent gap between viewers and non-viewers, while in the third it to increase the gap.

ssing the Adequacy of the Correction

nder favourable conditions a relatively high correlation can be eloped when the variable being studied is dependent upon an ty of some kind, though it may, as instanced in Table I, be as as + 0·5 or + 0·6 in the case of an attitude (e.g. attitude towards g to France). A multiple correlation as low as this may well be eflection of multi-dimensionality in the 'variable' being studied er than a failure to identify relevant matching variables; but in case, however high the multiple correlation, a check on the quacy of any matching combination is essential.

is typical of the method that this check is indirect, and this was case in the 'Bon Voyage' study. Had it been feasible—as obviously as not—to test the two groups before the broadcast and then to y the correction formula to the score of the non-viewing group, accurate estimate of the adequacy of the correction could have made, for obviously a fully effective correction formula would removed *all* pre-broadcast difference. Something approximating his type of check *was* feasible, however, and was made without fering with the research design. It was necessary to make the broadcast tests on two groups of people approximating to the two -broadcast groups as closely as expectations allowed. Full details e methods used to secure these groups are given later, and for the ent it will suffice to say that, like the people taking part in the -broadcast tests, their names were available from the records col-d in the department's Daily Survey of Listening and Viewing. *st be clearly pointed out, however, that once these pre-broadcast groups were they did not take part in any way in further tests.* Their one and only tion was to provide a check on the adequacy of the correction ulae, and the post-broadcast groups were made up of entirely rent people. It must also be noted that since the programme's

content could not, in advance of the broadcast, be known exactly pre-broadcast tests could be only approximations to the post-broa tests, though there was nothing to prevent them from being very approximations. The check through the pre-broadcast groups c only be applied, of course, after the correction formula had developed through one of the post-broadcast groups. The nece adjustments to pre-broadcast scores were then made and any resi differences noted. Results are set out in Table III, and from the will be seen that even in the case of the attitude assessment the r of the correction was to eliminate pre-broadcast differences alr completely. A high degree of confidence could therefore be entertai that the residual differences between the two post-broadcast gr (see Table II) did in fact represent 'effects'.

TABLE III

Adequacy of the stable correlates

(*showing matching corrections applied to the scores of two groups, viewers and non-vie tested* before *the broadcast of 'Bon Voyage'*)

Variable under Study	Non-viewers' Scores		Viewers' Sco
	Unadjusted Averages	Adjusted Averages	Unadjusted Averages
Knowledge of French words and phrases	4·43	4·26	4·26
Attitude on issues related to visiting France	6·77	6·66	6·60

The Securing of Subjects

Although a description of the method used to secure subjects is essential to the exposition of research design, it is necessary f balanced description of this particular application of it. The secu of subjects was, without a doubt, simplified and made inexpensiv the special facilities of Audience Research. At the same time t facilities are by no means essential and are replaceable by prelimi survey work.

(*a*) *The Post-broadcast Groups.* The department's daily surve listening and viewing provides the names and addresses of a sam of people who heard or saw programmes broadcast on the day p ceding the survey, and records also each person's age, sex and so group. This made it possible to identify two groups for post-broad testing: (i) a group which had seen the programme; (ii) a gr of non-viewers who had not seen the programme and who w approximately the same as the other group in terms of age, sex social group.

(*b*) *Pre-broadcast Groups.* 'Bon Voyage' occurred as a section regular afternoon programme called 'Leisure and Pleasure', so t

ames and addresses of potential viewers of 'Bon Voyage' were able for invitation to viewer meetings before the 'Bon Voyage' dcast. A group of non-viewers, similar in terms of age, sex and group, could also be recruited (for pre-testing) from the survey ds.

Other Applications of the Method

hough I developed this particular 'stable correlate' adaptation ne study of the effects of specific programmes, I have since used a study, using survey methods, of some of the long-term psycho-l effects of television,[3] and it appears to be readily applicable to sociological enquiries. Indeed, I should like to suggest this tation as a general research technique for use in those studies of ts' where circumstances preclude the making of measurements re the event.' Such circumstances are only too common and : (*a*) where pre-testing is likely to affect subsequent exposure to nfluencing conditions and the performance on the final test; (*b*) e pre-testing is in any case out of the question because either (i) is no knowing 'before the event' precisely who will be exposed ect to be exposed to the influencing conditions, or (ii) the effect studied has either occurred or begun to occur before the com-ement of the study (as in sociological enquiries into the effects ema-going, publicity campaigns, and so on). Such circumstances ot necessarily enforce an *after-only* design, but where they do, the ations of that design *as normally applied* are generally well recognised. very least that the proposed adaptation can do in such studies is remove some of that part of the difference (between groups) n is not an effect of exposure to the events under study and (ii) to de some indication of the degree to which that extraneous dif-ce has in fact been removed. This in itself seems to me to be a l advance.

is tempting, however, in the light of the 'Bon Voyage' study, to rther than this. At its best—that is, granted high-powered corre-and an efficient checking device—the adaptation could make ble a relatively unambiguous assessment of effects. A realistic oximation to this condition can never be guaranteed, of course, use in the end the origination of good correlates is dependent the individual research worker and upon the nature of the ble being studied. At the same time it is usually possible in socio-l enquiries to spend a little time beforehand in developing, g, selecting, and improving the proposed correlates, and this can ibute greatly to their final predictive power. Another difficulty t the less controlled are the nature of the influencing events and onditions of exposure to them, the less direct is likely to be the able device for checking the adequacy of the corrections made—at up to three or four independent checks may in fact have to be ed.

Granted all this, however (and the demands are not excessive proposed adaptation can make of an *after-only* design a mea achieving reasonably accurate assessments of effects without inter in any way with that essentially multi-determinant context wh likely to surround and to condition the normal operation of the under study.

REFERENCES

[1] Belson, W. A. (1956). 'Learning and attitude changes resulting from view television series "Bon Voyage",' *Brit. J. Educ. Psychol.*, **26**, 31.

[2] Belson, W. A. (1952). *The Comprehensibility of 'Topic for Tonight'*. Aud Research Department (BBC) Report, LR/52/1080.

[3] Belson, W. A. (1956). 'The effects of television upon the interests and initi of viewers' (study nearing completion).

[4] Garrett, H. E. (1947). *Statistics in Psychology and Education*. Longmans, Green Co., New York.

[5] Hovland, C. I., *et alia* (1949). *Experiments on Mass Communication*. Prin University Press.

[6] Wilkins, L. T. (1953). 'Some practical applications of recent developmen statistical theory', *The Incorporated Statistician*, **4**, 92.

MISSING VALUES IN EXPERIMENTS ANALYSED ON AUTOMATIC COMPUTERS

MICHAEL HEALY and MICHAEL WESTMACOTT
Rothamsted Experimental Station

Mr Healy and Mr Westmacott describe a general technique for dealing with observations missing from block experiments analysed on automatic computers. When more than one observation is missing the technique is simpler than other methods hitherto described and has been proved in practice to be satisfactorily fast. It is applicable to any analysis in which least-squares estimates are derived.

N automatic computers are used to analyse large numbers of iments, the problem of missing observations has to be faced. e time and effort will be expended on a small proportion of the unless the procedure for dealing with incomplete sets of results is anised. Although programmes of considerable generality for the sis of block experiments have been devised (see Tocher[5] and ey[4]), the complexity of such programmes and the shortage of ge space in most present-day machines often necessitate in prac- ne use of a number of different programmes for different classes ign. A general set of instructions for coping with missing values, can be tacked on to any programme that may be written, is fore desirable. It is clearly an advantage if the number of these ctions is kept as low as possible.

is well known, experiments with missing values can be simply sed by inserting fictitious values chosen so as to minimise the al sum of squares (Yates[6]). The analysis proceeds in the ordinary using the fictitious values, and the number of missing observations tracted from the total and residual degrees of freedom; unbiased ates of the error variance and treatment means are then obtained. the fictitious values are determined by minimising a sum of es, the appropriate computations can be carried out in any of the ent ways that have been evolved for least-squares problems. With machines, three techniques are in current use:

ne missing values may be replaced by algebraic symbols in terms which the residual sum of squares can be written down. Mini- sation of the resulting quadratic form leads to a set of simul- neous equations which can be solved by any convenient method.

hen only one observation is missing, the appropriate fictitious value n be determined from a formula (see, for example, Anderson[2])

which will, of course, depend on the nature of the experim design. If several observations are missing the formula c used iteratively. This latter method is equivalent to solvir equations in (1) by the Gauss-Seidel iteration technique.

3. An analysis of covariance can be carried out on a number of ps variates, each of which takes the value 1 for one of the missing and the value 0 elsewhere (Bartlett[3]). The fictitious values are given by the regression coefficients. This in effect provi mechanical method of writing down the equations in (1).

None of these techniques is particularly well adapted to auto computers. Technique (1) is ruled out as involving algebra rather arithmetic; (2) will require a different programme for every de and (3), though general enough, needs a programme for mu covariance analysis which will not commonly be available. further techniques, specifically designed for automatic computers, recently been suggested. Tocher[5] has shown for a wide class of de that a simple extension of the basic programme can be used to s the equations of method (1). These are solved, and a second app tion of the basic programme completes the analysis. Hartley,[4] fo case of a single missing value, suggests evaluating the residual su squares by means of the basic programme, giving three arbi equally spaced values to the missing unit. If these values are x and x_3, giving sums of squares Q_1, Q_2, and Q_3, the value of x mak a minimum is readily found to be

$$x_2 + \tfrac{1}{2}d(Q_1 - Q_3)/(Q_1 - 2Q_2 + Q_3)$$

where $d = x_2 - x_1 = x_3 - x_2$. If there is more than one missing the process must be used iteratively, as in method (2).

The purpose of this note is to put forward another technique experience has shown to have advantages over those described a This is based on the fact that the fictitious values are actuall expected values of the missing units derived from the correct squares estimates of the block, treatment, etc., constants, so tha residuals for these units after fitting the constants must all be zer the analysis of an experiment on an automatic computer it is un practice at Rothamsted to include the evaluation of the residu each plot—that is, the deviation from the general mean, correcte block, treatment, and other effects. The residual sum of squares is obtained directly as well as by subtraction in the analysis of varia and this provides a comprehensive check on the calculations. Furt more, it has been found amply worth while in several series of ex ments actually to print out the residuals in the randomised orde which they occurred in the original experiment, since this has ma simple to examine the causes of unduly high standard errors—fer trends, mechanical or bird damage on edge plots, uneven germina of certain varieties, etc.

residuals are thus provided by the basic programme. Suppose, hat we insert guessed values for the missing units. As soon as the als have been determined, we can subtract from each guessed its residual. This in effect replaces each guessed value by the ponding expected value derived from the analysis based on the d values. The whole process can then be repeated as often as ary. When the residuals for the missing units are small enough, quired fictitious values have been attained and the analysis can mpleted in the usual way. It is clear that the number of extra ctions required to carry out this process will be quite small, since, he residuals have been obtained, all that is necessary is one on for each missing unit and some form of test to find out whether ocess needs repeating. This can often be done conveniently by aring values of the residual sum of squares in successive cycles.

sideration of the successive values of the residual sum of squares to a simple proof of the convergence of the iteration. Suppose ne residual sum of squares, when the guessed values are used in alysis, is R_1 and that the residuals of the guessed values are a_1, b_1, When these residuals are replaced by zeros by the process de- d in the preceding paragraph the residual sum of squares be- $R'_1 = R_1 - a^2_1 - b^2_1 \ldots$, which cannot exceed R_1. If we arry out another ordinary analysis of variance using the corrected we get a new residual sum of squares R_2. This must be less than al to R'_1, since it is obtained in effect from a *least-squares* analysis corrected data. Hence $R_1 \geqslant R_2$ and similarly $R_2 \geqslant R_3$ and so th equality only if all the residuals of the missing values are zero. equence R_i therefore converges to the value of the residual sum ares given by the exact least-squares solution.

actual number of instructions required by the missing-plot mme will depend to a large extent on the amount of printing at is called for as a visual indication of the progress of the itera- It is convenient (though somewhat time-consuming) to print out ry cycle the values for the missing units and the residual sum of s. A few more orders are required to correct the residual degrees dom. As applied to a programme for analysing single-replicate 3^3 ments, the number of instructions on the computer at Rothamsted compared with some 800 in the main programme). An example ree missing units is given in the table, which shows the figures d by the machine at the successive cycles of iteration.

of Iteration		0	1	2	3	4	5	6
s for the g units	1	47·8	46·5	45·8	45·4	45·3	45·2	45·2
	2	51·4	49·7	49·0	48·7	48·6	48·6	48·6
	3	47·2	42·6	40·7	39·9	39·5	39·3	39·3
ual SS	..	173·78	137·93	131·57	130·45	130·23	130·22	130·22

The process described has the necessary generality and is deci simpler than Hartley's technique, but might be thought to be slow—algebraically it corresponds to the solution of the basic equ of method (1) by what Aitken[1] has termed *diagonal iteration*. In fa method will usually be almost as quick as Hartley's technique single missing value, and considerably quicker for several such v Hartley's method requires just three applications of the basic gramme to determine a single missing value, together with a fi application for the final analysis; when several units are missing, ever, the method requires three applications of the basic progra for each missing value in every cycle of iteration. Practical exper with randomised block experiments has shown that four or five c of the present technique usually give sufficient accuracy on al missing values, even when these amount to ten per cent or more c experimental material.

The technique described above is a good example of the new proach to well-known computing problems that is necessitated b advent of automatic computers. Entry to and exit from these mach and the writing of special programmes for particular situations more expensive in time than the repetition of standard analyses. analyses, moreover, may be carried out more thoroughly than is u or profitable, on desk machines. The mere evaluation of the resid though a quite trivial matter theoretically, is in practice hardly undertaken by hand because of the labour and tedium involved. that the residuals are to be calculated in any case, what appears a sight to be an inefficient method of computation turns out to be factorily fast.

We are indebted to Dr F. Yates for the proof of convergence to Dr H. O. Hartley for allowing us to see his paper in advan publication.

REFERENCES

[1] Aitken, A. C. (1950). 'On the iterative solution of a system of linear equa *Proc. Roy. Soc. Edinb.*, **63**, 52.

[2] Anderson, R. L. (1946). 'Missing plot techniques', *Biomet. Bull.*, **2**, 41.

[3] Bartlett, M. S. (1937). 'Some examples of statistical methods of resea agriculture and applied botany', *J. R. Statist. Soc. Suppl.*, **4**, 151.

[4] Hartley, H. O. (1956). 'Programming analysis of variance for general pu computers', *Biometrics*, **12**. (In press.)

[5] Tocher, K. D. (1952). 'The design and analysis of block experiments', *Statist. Soc.*, B, **14**, 45.

[6] Yates, F. (1933). 'The analysis of replicated experiments when the field r are incomplete', *Emp. J. Exp. Agric.*, **1**, 129.

A COMMENT ON THE CONSTRUCTION OF PRICE INDEX NUMBERS

K. S. BANERJEE

In an article on index numbers that appeared in this journal Dr Ruderman was led to consider the average price, obtained from family budgets, of a group of commodities. Dr Banerjee shows in the present article that, under certain conditions, an unbiased estimate of this quantity may be computed as the average price weighted with quantities sold in the market.

uction

a recent note[1] Ruderman drew attention to a point which is often cted in the construction of Cost of Living Index (CLI) numbers. lly, the mean-of-relatives form

$$\frac{\Sigma(p_1/p_0)p_0q_0}{\Sigma p_0q_0} \qquad \ldots\ldots(1)$$

ed in the construction of CLI numbers. Cancelling out p_0 of) with the p_0 of p_0q_0, (1) is reduced to the aggregative form of eyres, which is

$$\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \qquad \ldots\ldots(2)$$

rman points out that the transformation of (1) into (2) is not ble, because the p_0 elements in (p_1/p_0) and p_0q_0 are different. The p_1 and p_0 of the price relative (p_1/p_0) are market (store) prices, eas the values of p_0 implicit in the expenditure weights p_0q_0 are ned from family living studies. The latter price p_0 relates to a of commodities and will represent the average price of *all* goods class actually purchased at *all* prices actually paid in *all* markets nised. The two p_0 elements are therefore different and cannot, as cancel out. Using the subscript s to identify market (store) and f for prices obtained from family budget studies, the expres- or the weighted mean-of-relatives is obtained as

$$\frac{\Sigma(p_{s1}/p_{s0})p_{f0}q_0}{\Sigma p_{f0}q_0} \qquad \ldots\ldots(3)$$

ating the difference of p_{f0} from p_{s0} by some ratio a such that ap_{s0}, Ruderman calculates, for some commodities, the values of ch have been found to be different from unity.

Suggested Method of Computation of the Price Relative

In some places the prices p_1 and p_0 of the price relative (p_1/p weighted with quantities sold *in the market*. It will be shown in th that the average price p_0 weighted as such is an *unbiased estimate* under certain reasonable assumptions. If this be so, the deviatio from unity will be a matter of sampling fluctuation.

Let us explain with reference to an example what we mean b average price weighted with the quantities sold *in the market* and such average prices claim to meet the point raised by Ruderma the major group Food, 'leafy vegetables', for instance, will appe an item of consumption. Usually one weight will be available fo item 'leafy vegetables', which is the proportion of its expenditu total expenditure (or to expenditure on Food, if the constructio CLI is sought to be completed in the first stage individually for of the major groups, and in the second stage for all the major gr combined together).

There may be about, say, sixty varieties of 'leafy vegetables' sumed by the class of families for whom the CLI is constructed. expenditure p_0q_0 on this item, which is obtained from family bu studies, will be the total expenditure incurred on all the sixty vari

Where monthly indexes are required, the price reporting ag while collecting prices on several dates of the month from a markets (or a sample of markets) patronised, record also the vo of sales against the items priced. Usually the volume of sales (or sales) for the whole market day is recorded against the item. Coll of information on the quantities sold is possible in respect of nea the food items, including vegetables, fish, meat, eggs, fruits, etc. are then averaged with quantities sold as weights. It may be bo mind that here the prices are weighted with quantities sold *in the m* while the quantities q_0 of p_0q_0 are quantities consumed *in the j* There is therefore a difference between the two.

Comparison with Ruderman's Average Price

Let there be M varieties of the item under consideration (say, vegetables') and let there be N families from which budgets collected in the base period; also let q^0_{ij} ($\geqslant 0$) be the quantity sumed of the ith variety by the jth family. Then, placing the script 0 above for the sake of convenience,

$$p_0q_0 = \sum_i \sum_j p^0_i q^0_{ij} \qquad \ldots$$

Let us express p_0q_0 in the form $\bar{p}_0Q_0$ where $\bar{p}_0 = C_0/Q_0$, C_0 i right-hand side of (4) and $Q_0 = \sum_i \sum_j q^0_{ij}$. That is, $\bar{p}_0$ is the average weighted with quantities consumed in the family. If the price p_0q_0 could be expressed as a single price figure, it should be the as $\bar{p}_0$. This $\bar{p}_0$ is notationally equivalent to p_{r0} of Ruderman.

it were possible to collect the quotations of prices together with the ıtities purchased by the families in respect of all the M varieties in all the markets patronised by the class of people (say, the work-class in an industrial area in which the markets may be localised) ll the market days, it would have been possible to secure, for each ety, the total quantity consumed in the N familes. In that case, ıverage price weighted with the quantities sold *in the market* would been exactly the same as p_{f0}. But this is impracticable. Usually ection is made of m varieties out of M, and quotations of prices and ıtities are collected on some of the market days of the month (if thly indexes are constructed). Only a fraction of the N families probably make purchases on those market days.

et $\hat{Q}_i$ (dropping off the subscript 0) be the total quantity sold of ith variety, and $\hat{Q}$ the total quantity sold of all the m varieties. tiplying $\hat{Q}_i$ by $p^0{}_i$, we get $p^0{}_i\hat{Q}_i$, which is the total value of sales of ith variety. Written in full, $\hat{Q}_i$ will be equal to

$$q'_{i1} + q'_{i2} + \cdots + q'_{in_i}$$

re q'_{ik} $(i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n_i)$ is the quantity pur-ed of the ith variety by an individual family. Written in full, the ket transactions in respect of all the m varieties will look like

$$\begin{aligned}\sum_{i=1}^{m} p^0{}_i\hat{Q}_i = {} & p^0{}_1(q'_{11} + q'_{12} + \cdots + q'_{1n_1}) + \\ & p^0{}_2(q'_{21} + q'_{22} + \cdots + q'_{2n_2}) + \\ & \cdots\cdots\cdots\cdots \\ & p^0{}_i(q'_{i1} + q'_{i2} + \cdots + q'_{in_i}) + \\ & \cdots\cdots\cdots\cdots \\ & p^0{}_m(q'_{m1} + q'_{m2} + \cdots + q'_{mn_m})\end{aligned} \qquad \ldots\ldots(5)$$

ation (5) may be expressed in the form

$$\sum_{i=1}^{m} p^0{}_i\hat{Q}_i = \hat{p}_0\hat{Q} \qquad \ldots\ldots(6)$$

re $\hat{p}_0 = \hat{C}_0/\hat{Q}$, and $\hat{C}_0$ is the left-hand side of (5). $\hat{p}_0$ is therefore ıverage price weighted with the quantities sold.

ıe above quantities q'_{ik} $(i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n_i)$ will their equals in the quantities $q^0{}_{ij}$ $(i = 1, 2, \ldots, M;\ j = 1, 2, \ldots, N)$ of (4), but it is not known which of the former is equal to h of the latter.

ıy of the N families can have its transaction in respect of any of the ties on any of the market days. The occurrence of q'_{ik} can there-be taken as random.

ıce $p^0{}_i$ is constant over the ith variety and since the occurrence of s random,

$$p^0{}_iq'_{ik} = x_{ik} \quad (i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n_i)$$

may be taken as a random variable in the ith variety. Under assumptions, the sum total of the market transactions may be tak the result of a two-stage sampling where m varieties are chosen, i first stage, out of M, and n_i transactions are made in the ith va The unbiased estimate of (4) will then be given by

$$\frac{M}{m} \sum_i \frac{N}{n_i} \sum_k x_{ik} \qquad \ldots$$

The n_i's are not known. But it may not be unreasonable to ass equality of the n_i's, as, whenever a purchaser comes to the marke will make his complete purchase of the item; that is, he will purc from as many varieties as he may have preference for. Let $n_i =$ (7) will then reduce to

$$K \sum_i \sum_k x_{ik}$$

where $K = MN/mn$. That is,

$$C_0 = \mathrm{E}(K \sum_i \sum_k x_{ik}) = K\mathrm{E}(\hat{C}_0)$$

where E stands for mathematical expectation. In the same way it be shown that

$$Q_0 = \mathrm{E}(K \sum_i \sum_k q'_{ik}) = K\mathrm{E}(\hat{Q})$$

Now, if it could be assumed that

$$\mathrm{E}(\hat{p}_0) = \mathrm{E}\left(\frac{\hat{C}_0}{\hat{Q}}\right) = \frac{\mathrm{E}(\hat{C}_0)}{\mathrm{E}(\hat{Q})}$$

then

$$\mathrm{E}(\hat{p}_0) = \frac{C_0}{Q_0} = \bar{p}_0$$

that is, $\hat{p}_0$ is an unbiased estimate of $\bar{p}_0$, which is the same as I Ruderman.

REFERENCE

[1] Ruderman, A. P. (1954). 'A neglected point in the construction of price numbers', *Applied Statistics*, **3**, 44.

NOTES AND COMMENTS

many years biologists, chemists, physicists and engineers have had the it of abstracts covering substantially all the literature of their subjects, nly recently has anything of the sort been developed for applied ticians. The *International Journal of Abstracts on Statistical Methods in try*, issued by the International Statistical Institute, is in its third year; now is announced *Quality Control and Applied Statistics Abstracts*, published terscience Publishers Inc. of New York. This is an ambitious project, ing by its abstracts some 400 journals, each abstract being arranged now fully the significant contribution of each article so that the result be immediately useful, and it should be unnecessary to consult the nal article'. The fullness of the sample abstracts given in the prospectus doubtedly attractive to a prospective subscriber, but it also provokes tions on the dividing line between legitimate abstracting and plagiarism. many people are content not to consult original articles that originating als cease publication through lack of support, where will the abstracting als go for articles to abstract? The subscription price for the journal announced is $60 per annum.

★ ★ ★

e Institution of Engineering Inspection, which has covered the more lished methods of inspection since its foundation in 1919, has this year the step of altering its articles of association to include quality control ell. The definition of members now reads: 'Persons qualified . . . to rtake the full responsibility of inspection or quality control by statistical her recognised methods'. The description of various other grades of iation has been altered accordingly. An attempt will be made to gthen the institution by drawing members from a wider field and, by s of meetings, to direct the attention of managements in particular to possibilities of the newer techniques. The old-established journal *eering Inspection* is being replaced by a new journal: *The Inspection eer* (*Incorporating Quality Control*). The latter will pay particular attention se studies and practical applications. We welcome these ventures, with opportunities for a better liaison between the engineer and the tician and for stimulating a wider interest in statistical methods.

★ ★ ★

e of the projects set up by the European Productivity Agency, which art of the Organisation for Economic Cooperation in Europe, had as ject the sending of a study team from Europe to the US to examine ods of training in quality control methods currently in use there. A drawn from the ten or so countries participating in the project left on 26th April for a stay of about sixty days in the US, where they were the guests of the US Government through their Foreign Operations nistration. They were to have opportunities to attend a short course of

the type given to executives, to see in-plant training, and to see the t niques actually in use in a variety of industries. They were to stay enough to participate in the Annual Conference of the American Socie Quality Control, which was held this year in Montreal.

Mr C. S. Smith, head of the Mathematics Department of the Lei College of Technology, is the UK member of the team. Mr Smith numerous industrial contacts, especially in and around Leicester, a number of these interested firms have contributed towards the cost c journey. The Leicester College is already well known for the courses it both residential and part-time, in Industrial Statistics; the present pr should give them added stimulus. It is understood that Mr Smith wi willing to share his experience by undertaking a limited amount of lectu outside Leicester; in particular it is hoped to arrange, shortly after return, for him to address a meeting of those in this country especi interested.

★ ★ ★

The Industrial Applications Section of the Royal Statistical Society h a week-end conference on 'New Techniques and Applications of Stati in Industry' at Florence Nightingale Hall, University of Nottingham, f 20th to 23rd July. After Dr S. A. Wells, OBE, head of the Departmen Industrial Economics, had welcomed the conference on behalf of the U versity, Professor E. S. Pearson, CBE, President of the Royal Statis Society, in opening the proceedings, stressed the need for improving c munication between the mathematical statistician and his counterpa industry.

It was clear from the subsequent papers that the use of statistical met in improving productivity was rapidly increasing throughout indu Fields dealt with included production control and efficiency, accounta road transport operating, time study, machine interference, and sam and testing procedures. New techniques described ranged from si short-cut methods 'done on the backs of envelopes' to the use of elect computers for linear programming. Contributors to the discussions incl visitors from Holland, Czechoslovakia, and the USA, among them Prof J. W. Tukey of Princeton University.

The conference concluded with a discussion of methods of teac statistics within industry. Closely linked with this was a comprehe exhibition of models, charts, and other apparatus for use as teaching Visits were paid to the Boots pharmaceutical factories and to the Ho and Allied Trades Research Establishment.

★ ★ ★

We regret that the increased cost of printing will necessitate a small crease in the price of *Applied Statistics*. This increase will take effect v the first issue in 1957. Particulars of the new subscription rates are gi in the enclosed order form.

MEETINGS OF SECTIONS OF THE ROYAL STATISTICAL SOCIETY

INDUSTRIAL APPLICATIONS SECTION

BIRMINGHAM AND DISTRICT GROUP

Iarch 1956. Statistical Tests—Do the Assumptions Matter?: G. E. P.
x

pril 1956. A Short Review of some New Techniques: C. H. Leigh-gmore

Iay 1956. One-day Conference: Statistical Methods and Industrial periments

e object of the conference, held at the Birmingham College of Tech-y, was to enable scientists and technologists working in industry to int themselves with the benefits to be obtained from the design and is of experiments on statistical principles. Since it was addressed rily to non-statisticians, its approach was by way of practical illustra-of principles, hoping that in this way interest would be aroused.

P. F. R. Venables, Principal of the College, gave a short address of me to delegates. The first speaker was Dr H. C. Hamaker, of Philips rch Laboratories, Holland, who gave a lecture entitled 'Experimental ns in Action', in which the advantages of simple designs and a direct f presentation of results were stressed. Then followed Mr W. D. Ewan, ish Nylon Spinners, on 'Experiments in the Nylon Stocking Industry'. afternoon lectures had a common theme, that of the attainment of um conditions. Dr Box in his lecture 'Statistical Methods as an Aid chnical Development' showed how an existing process might be pro-ely improved by the recording and analysis of routine information on eepest ascent' principle. The last lecture, by Dr N. L. Franklin of the Atomic Energy Authority, on 'Optimum Conditions in a Chemical esis', was concerned with a special application of Dr Box's method.

conference was attended by over 300 delegates from many parts of untry, and to judge by the discussions, considerable interest was ated.

LONDON GROUP

pril 1956. Microspores and the Identification of Coal Seams: R. C. linson

ay 1956. The Sampling of Road Traffic: T. C. Tanner

MERSEYSIDE GROUP

arch 1956. The Contribution of Statistics to the Study of Road dent Causation: F. Garwood

10th April 1956. A Discussion on Sequential Methods: F. Downton

Mr Downton explained various types of sampling schemes and underlying assumptions. He illustrated his talk with an example sequential *F*-test, showing the full calculations and a shorter method on ranges. He also discussed two-sided sequential tests.

NORTH-EASTERN GROUP

14th March 1956. Statistical Applications of Automatic Computers: Page

SHEFFIELD GROUP

26th April 1956. Figures for the Manager: Stafford Beer

SOUTH WALES GROUP

21st March 1956. Annual General Meeting

This was followed by short papers and films. Dr K. W. Kemp dem strated the periodograph which he uses in the analysis of contin measurements to give both a qualitative and a quantitative picture o variations being examined. Mr A. M. Whitehouse explained the typ data that become available in the course of accident prevention campa He went on to discuss the validity of various comparisons of accident within and between factories. The titles of the films were: Frequ Curves, Techniques of Sampling, and Marketing.

STUDY SECTION

BRISTOL GROUP

15th March 1956. The Propensity to Save in Britain: L. R. Klein

Dr Klein told of the work of the Oxford Savings Survey in attempt chart the pattern of saving in various groups of the community. discussing the definitions of the basic quantities he gave examples of tions expressing saving in terms of them, and commented on the implic for economic policy.

MEDICAL SECTION

27th September 1955. A Study of the Physiological Effects of Hot Cli J. O. Irwin

25th October 1955. Family Sickness as reported by the Housewife Stein and S. Sklaroff

29th November 1955. From Graunt to Geneva; a History of Di Classification: A. H. T. Robb-Smith

31st January 1956. Physique; its Measurement and Influence: R. W. Pa (Joint meeting with the Study Section)

28th February 1956. Ageing in Industry: F. Le Gros Clark

27th March 1956. Statistical Problems of Surgical Follow-up: J. A Waterhouse

24th April 1956. Some Statistical Problems in Epidemiology: C. C. S

29th May 1956. Should Research be based on Routine Hospital Reco A. C. Dornhorst, J. H. Bignall, and M. P. Curwen

BOOK REVIEWS

ın Mannheim and Leslie T. Wilkins: *Prediction Methods in relation to Borstal Training.* ɔn: HMSO, 1955. Pp. viii + 276. 17s. 6d.

s book is the first of a series entitled 'Studies in the causes of delinquency ıe treatment of offenders', to be published for the Home Office by D. The authors are to be congratulated on their use of objective fic methods of investigation in a field where emotional prejudice and ported opinion are all too common.

situation considered by the authors is this. After release from a l institution some youths commit further crimes and some do not—the criterion of 'failure' or 'success'. The problem is to see how easily n discriminate between these two groups, using factual data appearing Borstal files and in other documents, and to compare results with tions based on the subjective assessments of Borstal authorities. A number of factors, relating to environmental conditions and past , were tested for association with success or failure. Fairly rough and methods were used for reducing the factors to a short list, which then ed, as the most relevant items: evidence of drunkenness, previous es, not living at home, living in industrial area, and shortness of s in any one job. Numerical scores were then derived by means of orm of discriminant analysis. This led finally to a graph or tables vhich the chance of success could be determined, given the total score ed by any individual. These prediction tables were then tested by ation to a new group of Borstal lads, and fairly good agreement en hypothesis and observation was obtained. It is interesting to note ıe information of predictive value related largely to the pre-Borstal of the youths. Little further accuracy was added by taking data on Borstal experience into account. Moreover, the statistical predictions on pre-Borstal histories were appreciably better than Governors' and masters' subjective prognoses on lads just about to be discharged from tal.

ough the idea of using prediction tables is not a new one, the present s a considerable advance on previous work, both in its use of powerful cal methods and in submitting the results to a proper validation test. ay is now open for further progress along similar lines.

ill be noticed that the above approach to social problems, whereby ve decisions are aided by the objective calculation of risks run in nt circumstances, is essentially the attitude of operational research—made by the authors themselves. When ultimately such prediction are available for other forms of treatment as well, e.g. prison, fines, ion, or discharge, the decisions of the Courts as to the disposal of rs should be materially improved. This is not to decry 'pure' research inology, which may well have an important role to play. However, orth noticing that, using the operational approach, much can be f immediate value to society without the development of abstract

The general arguments of the book are closely and carefully reasone difficulties and objections are fully discussed. Emphasis is placed practical aspects of the work rather than on the underlying mathen theory. While this is no doubt correct for a publication of this kind, be hoped that this theory will be presented in fuller detail elsewhere. outstanding point of importance is the need for a significance test f agreement of predictions with observations.

There is a great deal of useful and interesting information in this both on previous work and on the present research. One point special mention is the neat use of the new method of analysis to shov open Borstals have slightly better success rates than closed, even allowance has been made for differences in the quality of entries. An finding that may surprise some is the list of factors apparently not affe success or failure.

This book is certainly a landmark in the development of objective me in criminology, and it is strongly recommended to all those who interested both in this subject and in allied fields. NORMAN T. J. BA

E. C. Fieller, T. Lewis, and E. S. Pearson: *Correlated Random Normal Deviates* (Trac Computers No. XXVI, edited by E. S. Pearson, Department of Statistics, Univ College, London). Cambridge University Press, 1955. Pp. xv + 60. 10s. 6d.

These tables give in effect 3 000 random pairs from each of nine biv normal distributions with correlation coefficients 0·1 (0·1) 0·9. They been compiled from Wold's tables of random normal deviates with m and variance 1 (Tract No. XXV). The nine sets of pairs are form combining different variates x_1, x_2, etc., with a common variate x_0. variates are given to two places of decimals.

At the foot of each page the following summary values for the fift on that page are given: sums of variates, sums of squares of variates of products with x_0, and sample correlation coefficients. The last items are given to four places of decimals.

An introduction to the tables gives details of the method of constr employed and a description of the various tests of randomness ap together with the results obtained. The standard of reproduction tables is reasonable, and they should prove useful to all who have to co sampling experiments on correlated variables. D. G.

PUBLICATIONS RECEIVED

Norbert Lloyd Enrick: *Modern Mill Controls*. New York: Modern Textiles Magazin Rayon Publishing Corporation, 1956. Pp. 28. $2.

Cost Reduction. London: The Institute of Cost and Works Accountants, 1956. Pp. 4

INDEX

, *M. A.*, 66
nt statistics, 214
ting records, sampling problems, 86
atic computers, method of dealing
ith missing values, 203

N. T. J., 146, 216
e, K. S., 207
, M. K., 122
, M. S., 70, 141
D. G., 144, 216
W. A., 195
H. E., 67
E. P., 213
T., 66
R. R., 141

T., 143
es:
sh and American, 177
roduction and distribution, 177
e and Choice by Cardpack and
nessboard: Vol. 2', review, 68
an extension of the use of, 132
operations as statistical processes, 65
orothy E., 49
operations, applications of statistics
improve, 66
unication of Ideas, The', review, 143
ious measurements, analysis of, 139
butions of Survey Methods to Eco-
mics', review, 141
limits, modified, 12
ated Random Normal Deviates',
view, 216
tions, are they any guide to predictive
lue?, 113

m, J. S., 143
n, F., 214
rth, W. E., 65

an Productivity Agency, 211
W. D., 213
nental designs in action, 213

nalysis, for examining complex data
foodstuffs, 32
l experiment:
confounded, 122
ng of interaction in, 64
ne, R. A., 65
R., 113

Fieller, E. C., 216
Finney, D. J., 132
Fisher, Sir R. A., 68
Franklin, N. L., 213

Garwood, F., 66
Goobey, G. H. R., 1
Grant, E. L., 144
Gray, P. G., 73
Gregg, J. V., 65
Gridgeman, N. T., 106

Hamaker, H. C., 213
'Handbook of Industrial Engineering and Management', review, 144
Harper, R., 32
Healy, M., 203
Hendry, I. F., 65
Hill, D., 12
Hogben, L., 68
Holdsworth, E. S., 67
hospital planning and design, statistics in, 146
household income and expenditure, field work in sample surveys, 49

Kay, E., 65

'Industrial and Engineering Chemistry', 62
Industrial Applications Section of the Royal Statistical Society:
Birmingham group conference, 62, 213
conference, 212
industrial censuses, British and American, 177
industrial experiments, statistical methods and, 62, 213
'Inspection Engineer (Incorporating Quality Control), The', 211
Institution of Engineering Inspection, 211
interviewer variability, examples of, 73
'Introduction to Stochastic Processes, An', review, 70
inventory management, 139
investment of funds, use of statistics in, 1
Ireson, W. G., 144

James, S. F., 86
Jessop, W. N., 158
Jowett, G. H., 70, 138

Katona, G., 141
Kemp, K. W., 139, 214
Klein, L. R., 141, 214

Lansing, J. B., 141
Leander, E. K., 132
Lewis, T., 216
linear programming, 166
  in practice, 65
  the dual problem in, 140
logarithmic transformation, 63
Lomax, K. S., 139

Mackenzie, H. C., 140
Mannheim, H., 215
market developments, use of statistics in the study of, 65
market research data, application of statistics to, 67
Maurice, Rita J., 177
Mead, F. C., Jr., 122
medical care, assessing demand for, 146
missing values, dealt with by automatic computers, 203
Monte Carlo methods, applied to industrial problems, 158
Morgan, J. N., 141
Moroney, M. J., 140
Mosteller, F., 141

opinion survey, 66
optimum conditions in chemical synthesis, 213

Pearson, E. S., 145, 212, 216
percentiles, estimation of, 137
periodograph, 214
personnel selection, regression control chart for use in, 20
personnel statistics, 67
Plackett, R. L., 64, 68
'Population Statistics and their Compilation', review, 71
'Prediction Methods in relation to Borstal Training', review, 215
price index numbers, construction of, 207

'Quality Control and Applied Sta Abstracts', 211
quality control in the paper industry,
Quality Control Service for Industry land), 62

Rivett, B. H. P., 66
road accidents in Great Britain, 66

samples of two, total and differen 140
sample surveys, field work in, 49
Sandon, F., 20
saving in Britain, 214
scarce materials, allocation between ducts, 166
sequential methods, 214
Shankleman, E., 72
Smith, C. S., 212
snap observation techniques, 66
stable correlate method, 195
'Statistical Methods for Research Wor review, 68
statistical returns, use of, 156
'Stochastic Models for Learning', review

tasting experiment, 106
technical development, aided by stati methods, 213
television, studying the effects of, 195

Utting, J. E. G., 143

Westmacott, M., 203
Whitehouse, A. M., 214
Wilkins, L. T., 144, 215
Williams, N., 166
Wolfenden, H. H., 71

Youden, W. J., 62

Printed in Great Britain at
THE KYNOCH PRESS,
Birmingham